ELECTRONIC HEALTH RECORDS (EHR) QUALITY CONTROL AND TEMPORAL DATA ANALYSIS FOR CLINICAL DECISION SUPPORT

A Dissertation

Presented to

The Academic Faculty

by

Janani Venugopalan

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the

Wallace H Coulter Department of Biomedical Engineering

Georgia Institute of Technology and Emory University

December 2018

ELECTRONIC HEALTH RECORDS (EHR) QUALITY CONTROL AND TEMPORAL DATA ANALYSIS FOR CLINICAL DECISION SUPPORT

Approved by:

Dr. May D. Wang, Advisor
Wallace H Coulter Department of
Biomedical Engineering
Georgia Institute of Technology

Dr. Mark Braunstein
College of Computing
Georgia Institute of Technology

Dr. Robert Butera
Wallace H Coulter Department of
Biomedical Engineering
Georgia Institute of Technology

Dr. Kevin Maher
Department of Pediatrics
Emory University

Dr. Peng Qiu
Wallace H Coulter Department of
Biomedical Engineering
Georgia Institute of Technology

Dr. Nikhil Chanani
Department of Pediatrics
Emory University

Date Approved: [11, 08, 2018]

To my family and friends

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

GDP	Gross Domestic Product
HITECH	Health Information Technology for Economic and Clinical Health
ACA	Patient Protection and Affordable Care Act
ICU	Intensive Care Unit
MCAR	Missing Completely at Random
MAR	Missing at Random
MNAR	Missing Not at Random
NER	Not Easily Recoverable
EHR	Electronic Health Records
LSTM	Long Short Term Memory
NGRI	National Human Genome Research Institute
SNP	Single Nucleotide Proteins

SUMMARY

The United States of America has the highest spending in healthcare; however it ranks 37 in the quality of care. In addition, the cost of healthcare is increasing at a rate higher than that of the gross domestic product(GDP). This had led to some healthcare reforms such as the Patient Protection and Affordable Care Act (ACA) and meaningful use of electronic health records (EHR). The advent of "Big Data" era and meaningful use of EHR data has led to widespread use of patient data for clinical decision support. However, given the complex nature of the data, its volume and velocity, decision making is challenging.

The objective of this research is to develop methodologies for clinical decision support which target the prevention of readmission while reducing adverse events such as mortality, cardiac arrest and long stay in critical care units. We address challenges such as 1) missing data 2) the sequential nature of records in the ICU and 3) integration of heterogenous data for analysis. In this thesis, we developed novel strategies to solve these issues and contribute to this field of computer aided diagnosis using the three specific aims:

Specific Aim 1: To improve predictive performance by developing imputation techniques for missing data in EHR

Specific Aim 2: To develop predictive models for temporal EHR data

Specific Aim 3: Data Integration of EHR data using deep learning based predictive models

CHAPTER I

INTRODUCTION

1.1. Need for Clinical Decision Support Systems for Electronic Health Records

The United States of America (USA) has the highest total spending ($3.1 trillion in 2013 [2]) and per capita (>$8,000 USD) spending in healthcare[3]; however it ranks 37 in the quality of care [4] as measured using some key indices such as accessibility, equity, efficiency, care quality and health outcomes (infant mortality, life expectancy etc.)(Figures 1.1, 1.2). In addition, the cost of healthcare is increasing at a rate higher than that of the gross domestic product(GDP) [5]. Therefore it stands to reason that unless the process of healthcare is made more efficient, the country may no longer be able to sustain its aging population with the level of appropriate healthcare. This had led to some healthcare reforms such as the Health Information Technology for Economic and Clinical Health (HITECH) act in 2009 [6] and Patient Protection and Affordable Care Act (ACA), commonly called Obamacare, initiative in 2010 [7], to promote the "meaningful" use of electronic health record systems (EHR) [8]. These initiatives, amongst other things, propose the use of data-

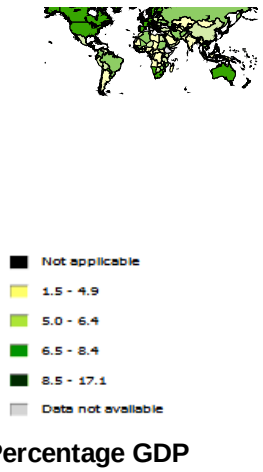


Figure 1.1: Average per capita health spending 2014 Figure from WHO Global Health Expenditure Database

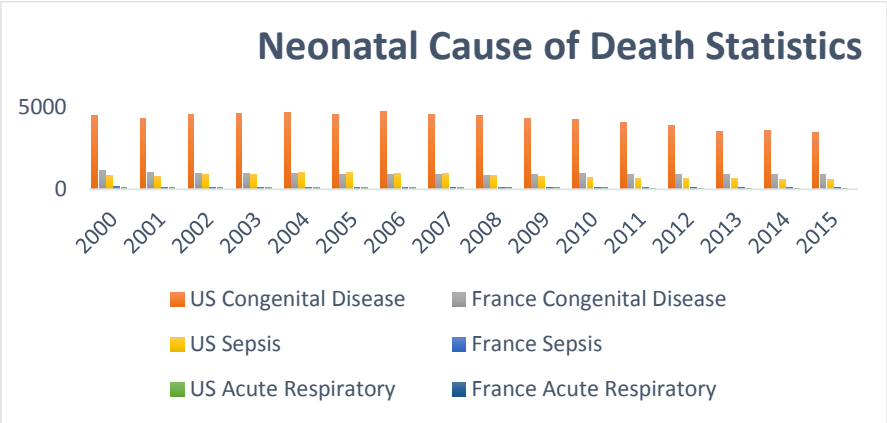


Figure 1.2: Neonatal cause of death statistics, Data from WHO Global Health Expenditure Database

driven methods for improving the quality of healthcare while decreasing the cost incurred. As the first few steps towards the "meaningful" use of EHR systems, it became mandatory for hospitals and vendors to support electronic archival of health records and support data interchange policies and APIs [8, 9]. This has resulted in the rapid growth of healthcare data volume, which was 150 Exabytes (150×10^{18} bytes) in 2011 [10]. In order to utilize this vast and complex source of data for clinical tasks such as clinical tasks: screening, diagnosis, prognosis, and resource management, Big Data analytics and tools are necessary.

1.2. Clinical Decision Support and Big Data Analytics in EHR

The concept of Big Data refers to the development of decision support systems which are capable of handling large and complex datasets[11] and assist the clinicians in their decision making. This process involves the processing of raw data to extract meaningful information, which in turn is modeled to obtain actionable knowledge using data mining techniques (Figure 1.3). Despite extensive research in the field, there remain challenges to the decision-making process. This is mainly due to the complexity of the data

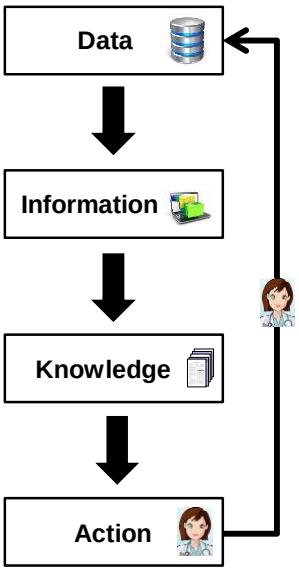


Figure 1.3: Big Data Analytics Pipeline in EHR data

sets involved. The complexity of the dataset refers to the "five Vs" of big data including variety, velocity, veracity, and value [12, 13]. Variety refers to the heterogeneous nature of data in most EHR systems, velocity refers to data acquisition rate, veracity refers to the quality of data and measured features, and value is the utility of the decision obtained from the data. The challenges posed by these can be referred to in data mining terms as

1) Data quality control
2) Irregular Sampling of Temporal Data
3) Heterogeneous data sources.

In this work, we use intensive care unit (ICU) data as a case study to develop and evaluate methods to overcome these challenges for clinical decision support.

1.3. Clinical Decision Support in ICU

More than 5 million patients are admitted annually to ICUs in the United States [14], with children and adolescents accounting for 18% of the hospital stays [15]. Children admitted to the ICUs require extensive medical care with high resource utilization [16] .

Within hospitalizations, the ICU is often the site of greatest resource utilization because these children require high-level nursing, frequent laboratory monitoring, and multiple invasive and non-invasive procedures – all of which come at a great expense and often with associated complications [17]. Length-of-stay has been shown to be a high-risk factor of long-term adverse effects (e.g. neuro-developmental disorders for young kids) associating with a higher cost of healthcare [18]. Most ICU studies have focused on adult populations (18 years and older) [19, 20], and have shown that the length of stay has significant contribution to increased life-threatening outcomes [21, 22] such as severe infection [23, 24], cardiac arrest [25], ventilation length [26, 27], one-year mortality [28, 29], ICU/hospital readmissions [30], acute kidney injury [21, 29], and hypotension [31-33] etc. Any intervention, pharmacologic or procedural, that could abbreviate their length of stay in ICU and prevent readmissions would have a significant impact on the child's quality of life and society's overall resource utilization [34-36].

Research investigating the causes of adverse events in the pediatric populations is relatively rare. Although some scores of ICU risk assessment are available for pediatric populations (e.g., pediatric index of mortality (PIM1&2) [37], and pediatric risk of mortality (PRISM, [38-41])), most of them target population of 2-16 years and they represent correlations without accounting for the inherent temporal changes embedded in the electronic health record data. Scores such as PRISM and PIMS only handle general scenarios without the ability for disease-specific modeling, or assessment of ICU length-of-stay's impact on young children's long-term development. Studies have shown that data-driven methodologies outperform expert system based risk scores for predicting adverse events including mortality [28]. However, the ability to use these modeling

techniques for predicting adverse events in the ICU have been confounded due to inconsistencies in the data, the heterogeneity of data and the temporal nature of data. In this dissertation, we develop and evaluate our models on both adult and pediatric data.

1.3.1 Data Quality challenges in ICU data

The intensive care unit (ICU) is equipped with a multitude of monitoring and therapeutic equipment which generate large amounts of complex multimodal data [42]. This diversity and the number of parameters being monitored in an ICU make the resulting databases highly susceptible to several quality issues in data such as a) missing data and b) erroneous data entry [43-48]. The major reason for this situation is that despite comprehensive record-keeping, not all values or parameters are obtained in every case because tests or measures are only recorded when the clinical team suspects a clinical condition based on observations. Therefore, important events may be unobserved and no specific data is guaranteed at any time point. The presence of poor quality data in the database adversely affects the downstream processing and predictive modeling. Hence the issues of data quality pose significant challenges to decision support systems.

Errors & Inconsistencies in ICU Data

Errors and inconsistencies in clinical data are a challenge which has been widely recognized in the community. However, there is no consensus on the handling of these errors and inconsistencies [49, 50]. Errors in medical data detected using double entry method ranged from 2.6-26% [51]. They are generated due to systematic and human errors. Not all of the erroneous data is at random [51]. It is mainly detected by checking for impossible units and checking for the clinical limits for the data entry [52]. In some studies, erroneous data is checked using data distributions [53]. Inconsistencies such as wrong

spelling and typing errors are corrected, but other types of errors do not have a clear consensus for handling and may contribute towards poor decisions.

Missing Data in the ICU

Missing data is a significant issue in EHR analysis [54], with major repercussions on the validity of results generated from downstream data mining of this data. Conventional interpolation and imputation schemes generally perform poorly because there are no models for modeling the processes which generate missing data [55]. Current imputation models in the ICU either try to fill in missing data or delete records with missing values [56]. Filling in of missing values is done by using population averages or the means or median of the database [57]. Deletion of records is either list wise or pairwise deletion [58-60]. Deletion leads to a loss of statistical power and mean filling introduces errors in the data, which may not accurately reflect the underlying disease state [61]. More accepted technique for handling missing data are based on interpolation and model-based approaches such as interpolation, multiple imputation [47, 57, 61-64], expectation maximization [65-71], maximum likelihood methods [72-74] and hot–deck imputation [75, 76]. These models, though superior to conventional approaches and interpolation based methods still do not account of the patterns inherent in missing data or the missingness mechanisms. In statistics and financial literature, missing data is divided into three groups on the basis of missingness mechanism. They are (a) missing completely at random (MCAR), (b) missing at random (MAR) and (c) missing not at random (MNAR). Since there is a semantic gap in the actual terminology and its context in the ICU, we re-phrase these terms as follows: “Neglectable” also known as MCAR, “Recoverable” also known as MAR and “Not-Easily-Recoverable (NER)” also known as MNAR (Table 1.1).

Data is classified as “Neglectable” if the probability of missing data does not depend either on the missing values or other observed data. “Neglectable” data can occur in any clinical variable and is independent of observed data. Missing data is classified as “Recoverable” if the probability of missing data depends on the observed values of other features in the dataset. Missing data is classified as “NER” if the probability of missing data depends on the actual missing values.

Most current research in health which do use model-based approaches for handling missing data assume that all the missing data in the database is either “Neglectable” or “Recoverable” [47, 77]. Often times, these assumptions are not mathematically tested and may lead to biased conclusions [47, 63]. Some studies, such as those of Sun et al. [63] performed analyses to find out the effect of performing missing data analyses on the end results (effectiveness of lymphadenectomy). The methods compared were 1) multiple imputations to impute missing values; 2) deletion of cases with missing values; and 3) making cases with missing values a subcategory. However, this study also mentions that the missing data was more likely in patients who had a high tumor grade. This means that the data was likely MNAR and the validity of the assumptions made are questionable.

However, most of these analysis methods in the ICU which perform some operations on missing data, assume that all data is “Neglectable” or “Recoverable” and perform imputations [62]. This leads to bias in the results. Zelnick et al. [47] performed

Table 1.1: Missing Data Types Data Dictionary

Definition ICU Context	**Statistical Literature**	**Abbreviations**
“Neglectable”	Missing Completely at Random	MCAR
“Recoverable”	Missing at Random	MAR
“Not Easily Recoverable (NER)”	Missing Not at Random	MNAR

studies on traumatic brain injury to assess the changes in functional outcomes in TBI over long term versus those obtained over the short term. They correlated these features with long-term prognosis. Missing data was one of the major challenges encountered by this group. The used multiple imputation to impute missing data prior to prognosis prediction. However, they assessed the data only for "Neglectable" condition and they did not test the data for "NER" prior to imputation. The "Neglectable" analysis performed was also not a quantitative analysis. These can actually lead to biased conclusions, given that the paper could prove that the populations with higher missing data were distinct from the populations whose data was available.

Jing Tian et al [57] demonstrated the use of a novel imputation method which uses a combination of clustering and multiple imputation methods for imputing missing data in aerospace research. Their premise was that data imputed from most similar elements will better represent the data than ones which use all the data. They also proposed a Gray similarity metric to assign clusters to partial data. However, this approach makes use of only the complete data for cluster generation. This may not be optimal for ICU where most of the data have at least some fields missing.

Sun et al [63] performed analyses to find out the effect of performing missing data analyses on the end results, which was the effectiveness of lymphadenectomy in this case. The compared different three missing data handling techniques methods prior to estimating survival using multivariate cox regression. The methods were 1) multiple imputations to impute missing values; 2) deletion of cases with missing values, and 3) making cases with missing values a subcategory. Their results suggested that lymphadenectomy had no effects on survival. Previous studies which used multiple imputations reported the effect due to

effects of data handling. However, this analysis deletes data where the missing value is in fields other than tumor grade, these could lead to bias in this analysis. In addition, the study mentions that the missing data was more likely in patients who had a high tumor grade. This means that the data was likely "NER". These could have contributed to the changes in the multiple imputations.

Jenkins et al [77] measured the variability of direct nursing cost for similar patients and examined the characteristics of nurses assigned to different patient types. Their results showed a high variability in nursing intensity and cost per day amongst the different patient groups. In their analysis, however, though they analyzed the patterns of missing data, they did not perform significant steps to address the same. They deleted records with missing nursing data. This value amounted to approximately 25% of their sample size. This could have led to loss of statistical power and biased results

In this dissertation, we address this issue by developing novel imputation methods of each of the three categories.

1.3.2 Temporal Data Analytics using ICU Data

Conventional Data Mining Models with ICU Data

Population-based studies have shown that prolonged hospital stay contributes significantly to increased life-threatening outcomes, which are strongly associated with adverse events such as risks in the ICU environment [21, 22], severe infection [23, 24], cardiac arrest [25], prolonged duration of mechanical ventilation [26, 27], ICU/hospital readmissions [30], high red cell distribution width [56], acute kidney injury [21, 29], and hypotension [31-33]. Amongst these models logistic regression and Cox regression are the most common models used in the analysis of EHR data. They are advantageous since they

do not make any assumptions about the distribution of the data. Fuchs et al. used logistic regression and cox regression models to estimate the effect of age and disease severity on short- and long-term survival. Their findings suggested that age is nonlinearly associated with mortality and should be treated as an independent factor affecting mortality [78]. Lee et al. created regression models for patients suffering from hypotensive episodes with fluid levels and the dosage of vasoactive agents as predictor variables and mortality and the length of stay as response variables. They found that fluid resuscitation is beneficial for reducing hospital length of stay and the use of vasoactive agents increases in-hospital mortality [33]. The use of more sophisticated data mining techniques such as artificial neural networks, support vector machines, and decision trees, outperformed logistic regression based scores such as APACHE III with fewer variables [79]. Wong et al. used a back-propagation artificial neural network (ANN) and compared it with APACHE II scores and found that the ANN models outperformed the APACHE scores. They used intensive care unit data for approximately 8000 patients from 26 ICUs [80]. This study used that data which showed maximum deviation from mean values for predicting mortality. Nguyen et al. used k-nearest neighbors and decision trees on claims data from 120,000 people to assess the number of days of hospitalizations. They proved that a combination of regression and decision trees performed better than either method used separately [81]. This goes to suggest that nonlinear decision boundaries may help predict health data in a better manner.

Temporal Data Mining Models with ICU Data

Conventional models though very useful for finding risk factors associated with a specific disease, could not be used for predicting the risk of a specific treatment plan for

an individual patient and do not incorporate the temporal nature of the clinical data. Because there is often a delay between the occurrence of a treatment and its influence on the outcomes, it is important to study the temporal relationships between events that can provide evidence for clinical decision support. For example, myocardial infarction can be predicted based on the temporal changes in ECG, and, similarly, hepatitis can be diagnosed using the temporal relationship of viral counts.

The temporal models commonly seen in the literature include models such as sequence analysis [82-86], association rule mining [85, 87, 88], temporal Cox regression [89-91] and clustering [92, 93]. Sequence analysis and association rule mining based studies require extensive user input for identifying specific features whose patterns of correlation can be studied with respect to the target variable. In addition, they are not amenable for discerning relationships and patterns contributing to adverse events, from a large number of features. Regression [94] and clustering [95] based studies use information within a specific time interval for analysis. These studies do not account for the differing length of available data for different patients. Cox regression also does not account for the dependency between the consecutive time points. Graphical methods by Liu *et. al.* uses Gaussian processes (GPs) for time-series analysis [96]. Their assumption is that the data is piecewise linear and use only the GP coefficients for classification. Such models make the assumptions that ICU data can be approximated using piece-wise GPs. Stiglic *et. al.* used past recordings for a single patient to make predictions about a future time instant using LASSO regression [97]. The parameters of these models, are trained for each individual patients and do not make use of the information which can be learned from large databases consisting of multiple patients. Such models not only require the user to train the model for

each patient but also tend to over-fit the data. In addition, these models do not tell the clinicians if the patients are improving over time. To the best of our knowledge, there is only one study by Lin *et. al.* [98] which generates individual survival curves. They use a series of logistic regression models to calculate the hazard at each time instant. This approach is not only very computationally intensive but also does not account for the variation in the duration of ICU data.

Graphical sequential models such as Markov models and conditional random fields have been used for waveform analysis in EHR data [99-103] and may be adapted for other types of EHR data as well. Penny et al. used AR-HMM model to look at state changes in sections of EEG, recorded over the primary motor cortex, corresponding to imagined finger movements [104]. HMMs are generic probabilistic models whose observations may be of arbitrary complexity and may be generated via another, nested, probabilistic model [105]. These methodologies have a potential in sequential data found in EHR also.

1.3.3 Data Integration

Integration of data can be done at multiple levels a) heterogeneous data integration within EHR/ICU systems b) multi-modal data integration (e.g. EHR + genomic data integration)

Heterogeneous Data Integration at Multiple Temporal Scales

Clinical EHR data consists of administrative data (billing, insurance, procedures performed etc.), ancillary clinical data (vital signs, lab tests, medication etc.) , and clinical text (physician notes, text observation) [106]. All these values are collected at different stages by a variety of personnel as varying frequencies. As a result, the collected data is

multivariate and heterogeneous in nature[107]. The data consists of multiple types of events with some numeric values and some categorical values. In addition, the sampling is also irregular and at diverse temporal scales [108-110]. The advent of deep learning techniques such as auto encoders [111-114], convolutional neural networks [115, 116] and restricted Boltzmann machines [117-119] have shown progress towards addressing some of these challenges of data heterogeneity. The temporal aspects of the EHR data have been addressed using methods such recurrent neural networks [120], conditional restricted Boltzmann machines [118, 121] and long-short memory networks (LSTM) [122, 123]. These models have been used to predictions of a variety of disease conditions using clinical EHR/ICU data. Similarly, such methods have been used for waveform data as well for adverse event prediction [124-127]. However, the combination of temporal data from multiple temporal scales using such deep-learning methods is still is an open research area. In our work, we address this challenge using LSTM networks.

Multi-Modal Data Integration

EHR data analytics mainly focuses on predicting future health-related outcomes by leveraging personalized longitudinal data to support clinical decision making. However, the current data mining models which use only EHR data can predict outcomes only after the symptoms are seen. Big data analytics which combines genomic with EHR has been identified as one of the research directions which can be employed to improve the early prediction of diseases which have a genetic component [106, 128-130]. The potential for combining EHR data with genetic data has been shown by the eMERGE consortium [131]. The eMERGE network is a National Human Genome Research Institute (NHGRI)-funded consortium. The eMERGE network aims to identify causal genomic mutations (mostly

single nucleotide proteins (SNPs)) for phenotypic information (e.g., observable phenotypes for genetic disorders, drug responses, childhood obesity, and childhood autism) recorded in the EMR system, and then integrate identified genotype-phenotype associations into the system [131]. In addition, there is also some data mining efforts to show EHR and genetic data integration using techniques such as ontology mining with genotype mapping and frequent pattern analysis[132, 133]. However, despite the superior performance of deep-learning techniques, multi-modal deep learning analysis in healthcare seems to be an open research area.

1.4. Proposed Study and Organization of Dissertation

In summary, the major challenges of decision support systems are 1) records with missing data 2) the sequential nature data in the EHR and 3) heterogeneous and multimodal nature of data. In this dissertation, we aim to develop novel strategies to address these issues and contribute to this field of computer aided diagnosis through the development of robust models which are capable of analyzing time series data in the electronic health records (EHR). More specifically the three aims are as follows (Figure 1.4).

Specific Aim 1: To improve predictive performance by developing imputation techniques for missing data in EHR

Specific Aim 2: To develop predictive models for temporal EHR data

Specific Aim 3: Data Integration of EHR data using deep learning based predictive models

This dissertation focuses on predicting future health-related outcomes by utilizing longitudinal data, starting from improving the quality and integrity of raw EHR data, applying data mining and machine learning techniques with selected features to construct predictive models and finally delivering knowledge generated by predictive models to

support clinical decision making. Lastly, integrative analytics aims to establish a common platform that the highly heterogeneous biomedical data can be linked, modeled, and interpreted together.

The chapter 2 of this thesis focusses on the development of novel imputation techniques for missing data with a focus on the type of missing data. The chapters 3 and 4 deal with the development of temporal data mining models and the combination of temporal data at different time scales. Chapter 5 discusses the development of deep learning models for EHR data for integration of heterogeneous temporal data at different time scales, and chapter 6 discusses the use of deep learning models for integrating EHR and SNP data.

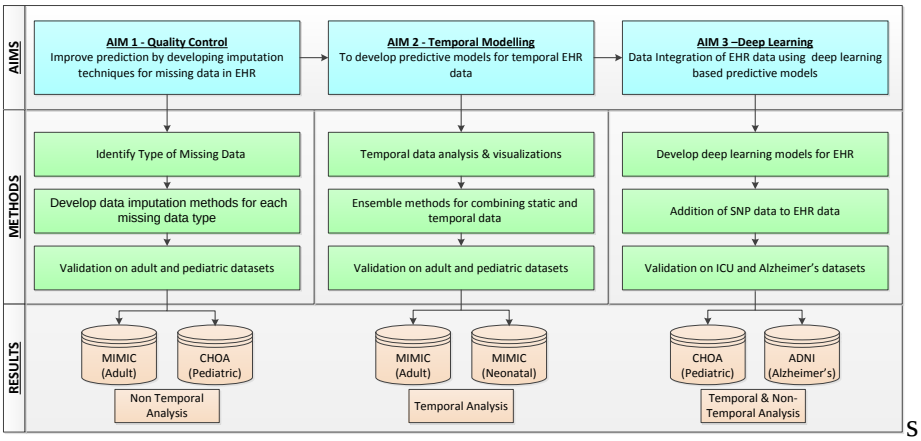


Figure 1.4: Specific Aims of this Dissertation

CHAPTER II

DATA IMPUTATION FOR MULTIPLE TYPES OF MISSING DATA

2.1. Introduction

As mentioned above the first objective of this dissertation is to address the issue of missing data and develop novel imputation techniques for each type of missing data. The databases containing ICU are highly susceptible to quality issues, such as missing information and erroneous data entry, which adversely affect the downstream processing and predictive modeling. Conventional missing data interpolation and imputation techniques perform poorly because there are no standards for modeling the missing data. Current models for imputing missing data include multiple imputation [47, 57, 61-64], expectation maximization [65-71], and hot – deck imputation [75, 76] techniques. These techniques are capable of handling only some types of missing data and hence lead to biased results[73] if used on all types of missing data.

In our research we categorize missing data into three the different types of missing data mentioned in Chapter 1, "Neglectable", "Recoverable" and "Not-Easily-Recoverable (NER)". Then, we address the issues of imputing "Recoverable" and "NER" data in the ICU by extending the clustering based approach of Tien *et. al.* [57] for "Recoverable" data imputation and developing a copula-based "NER" imputation technique. Our novel "Recoverable" imputation combines the benefit of both expectation maximization (accounts for distribution) and hot deck techniques (fewer effects due to cross user inconsistencies [134]). The novel imputation methods were then evaluated using two case studies. 1) Adult ICU data, and 2) Pediatric ICU data.

Table 2.1: Missing Data Types

Missing Data	Term in	Definitions	Examples
Neglectable	Missing completely at random	Missing data independent of missing values or other features	Data entry operator missing an entry, sensors falling off in measurement
Recoverable	Missing at random	Missing data dependent on other features but independent of the missing values	Inferring hematocrit from hemoglobin; or not assaying troponin for all patients with chest pain
Not-Easily-Recoverable	Missing not at random	Missing data dependent on missing values	All the responders to a drug not answering in a drug survey

2.2. Types of Missing Data

Missing data is classified as "Neglectable" if the probability of missing data does not depend either on the missing values or other observed data. "Neglectable" data can occur in any clinical variable and is independent of observed data. Missing data is classified as "Recoverable" if the probability of missing data depends on the observed values of other features in the dataset. Missing data is classified as "NER" if the probability of missing data depends on the actual missing values. In other words, given dataset X with missing data in feature $Y = [Y_{obs}, Y_{miss}]$, where Y is composed of both observed data Y_{obs} and missing data Y_{miss}, the data is "Neglectable" if $Y_{miss} \perp X$ and $Y_{miss} \perp Y$. The data is "Recoverable" if $Y_{miss} \perp Y$ but $Y_{miss} \intercal X$ and it is "NER" if $Y_{miss} \intercal Y$. Then we develop novel imputation methods for each type of missing data (Table 2.1).

2.3. Methods

2.3.2 Identifying the Type of Missing Data

As mentioned above there are three types of missing data (Table 2.1), "Neglectable," "Recoverable" and "NER." First, we analyzed missing data to find if they are "Neglectable". If not, then we distinguish between "Recoverable" and "NER". If the

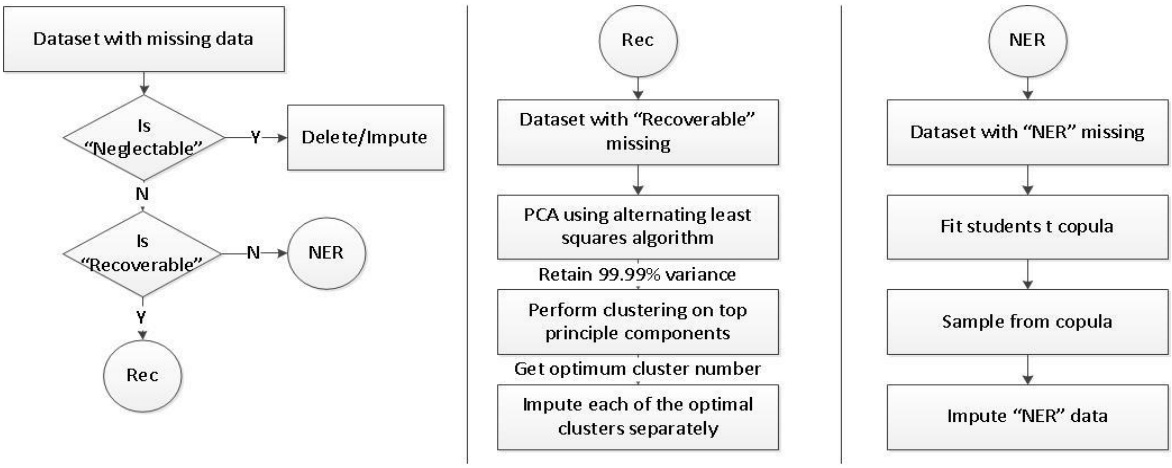


Figure 2.1: Imputing "Recoverable" data using clustering method & "NER" data from student's copula

data is "Recoverable" then we impute the data under "Recoverable" assumptions, else we estimate under "NER" assumptions (Figure. 2.1):

Test for "Neglectable".

There are two tests commonly used in literature, a series of t-tests [135] and Little's test [136], which test whether the missing data is "Neglectable".

A series of t-tests are performed to determine whether data is "Neglectable". By definition, "Neglectable" data refers to missing data that does not depend either on the missing values or observed data. For each feature, we divide the remaining features into two groups. The first group has the data where the test feature contains missing values, and the second group contains data where there is no missing value in the test feature. If the results of t-test show that the two groups are sampled from the same population, then the data is "Neglectable". Since the test is performed on a single feature at a time, missing data in other features was disregarded. We performed the t-tests with respect to each of the 'f' features with Bonferroni corrections to account for multiple testing at a statistical significance of 0.05. However, for a total of 'f' features, this required a total of 'f (f-1)'

comparisons where ‘f’ is the number of features, which can become computationally expensive for large datasets with thousands of features.

This issue is solved using Little’s test [136], which produces a score called Little’s score, used for distinguishing “Neglectable” missing data. Little’s score is obtained by comparing the means of the original data with maximum likelihood imputed data. This score follows the chi-square distribution if the data is “Neglectable.” [136]. A p value less than 0.05 rejects the hypothesis that the missing data is “Neglectable”. This was implemented in IBM SPSS. Due to memory constraints, we implemented this by taking features in batches to test for “Neglectable”. When a current batch was not “Neglectable” we did not perform further analysis on the batch. However, if any combination was “Neglectable” we combined it with more feature sets (adjacent set) and tested again.

Following the test for “Neglectable”, we then distinguished the “Recoverable” from “NER” data, prior to imputation.

Distinguishing “Not-Easily-Recoverable” from “Recoverable.”

Data is classified as “Recoverable” if the missing data depends on the other features, and it is classified as “NER” if the missing data depends on the missing values. Previous research suggests the use of classification schemes to distinguish “Recoverable” data from “NER” data [137, 138]. Cismondi et al. [138] used fuzzy classification schemes to distinguish “Recoverable” from “NER” data. They proved that non-imputation of “NER” data gives better results and lower bias compared to the imputation of all the values. The labels for training and classification was generated for each feature by assuming the value of 1 if data was missing and 0 otherwise. Any data that was missing and was labeled accurately was considered to be missing (“NER”), and those which were mislabeled were

considered to be imputable ("Recoverable"). This procedure was repeated for each of the different features. We report the correlation between the values of "Recoverable" and "NER" data evaluated using the different classification techniques. Then following this, we impute the "Recoverable" data and estimate the "NER" data using our novel methods described below.

2.3.2 Missing Data Imputation

Imputation of "Recoverable" Data

Imputation of data under "Recoverable" assumptions has been performed widely in medical literature using expectation maximization, and multiple imputations. However, these techniques tend to cause cross user inconsistencies and errors due to parameter estimations, which are avoided using a clustering based imputation [134]. Previous studies using clustering for imputation use only those records where all the data is available, for creating the clusters[57]. This approach is not very feasible in environments with a high rate of missing data such the ICU. Hence in our study, we propose an alternating least squares PCA based clustering approach before imputation so that the effect of missing data on the clustering is reduced.

Robust Clustering based Imputation

First, PCA using alternating least squares algorithm is performed on the data. Alternating least square based PCA can account for missing data while filling in the missing values for the principal components [139]. The total number of principal components were chosen to account for 99.99% (chosen to preserve most of the information content) of the variance. Then the top principal components were clustered

using k-means and fuzzy-C-means (fcm) clustering. In this study, we chose k-means and fcm as representative hard and soft clustering techniques[140].We used Calinski Harabasz, Davies-Bouldin and silhouette quality metrics for estimating the optimal cluster number. The optimal cluster number for each score was computed using the mean of five repetitions, and a voting principle was used to compute the number of clusters used for clustering ICU data. This step ensures that robust clustering can be performed even in the presence of noisy and missing data. Then the imputation was performed in each of the clusters. We limited the number of clusters to range from 2-20 in order to ensure data characteristics are captured, while keeping in mind the limitations of data size and computational cost. For imputing each of the clusters in this study, we used expectation maximization (Figure. 2.1).

Estimation of data under "Not-Easily-Recoverable" Assumptions

By definition "NER" data depends on the missing data and the patterns of missing data. The methods for dealing with this type of data in statistics are selection models, pattern mixture models [141] and drawn indicator models [142]. All these models assume a multivariate normal distribution for estimating "NER" data. A major issue with such models is that in the ICU scenario, the distributions are rarely normal.

In this analysis, we extend drawn indicator models for ICU EHR data. Drawn indicator models, use multivariate normal distributions which account for the "missingness" patterns in addition to the relationship between the features to model the missing data (Here, the "missingness" pattern is defined as the distribution of missing data where a value of 1 is given when a specific data is missing and 0 otherwise). When features are not normally distributed, then multivariate normal distributions become unreliable models for imputation [143]. We overcome this issue using copula functions.

A copula function couples N univariate marginal distributions together to form a joint distribution function of N standard uniform random variables. It has been shown to be invariant to elliptical distributions, deviations from normality and overcomes the issues of normal distributions. Hence, we fit a multivariate copula under "NER" assumptions to sample from for estimating the "NER" data. In our study, we use a t-copula which is a function of the features and the "missingness" pattern R, (defined as 0 when a certain data is observed and 1 when otherwise).

A function C : [0, 1] p → [0, 1] is a p-dimensional copula if it satisfies the following properties:

1. For all $u_i \in [0, 1]$, $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$.

2. For all u ∈ [0, 1] p (i.e. the dim , $C(u_1, \ldots, u_d) = 0$ if at least one of the coordinates, u_i , equals zero.

3. C is grounded and p-increasing, i.e., the C−measure of every box whose vertices lie in [0, 1] p is non-negative.

Each of the u is a the marginal distributions of the random variables.

Consider p continuous random variables ($X_1, \ldots, X_p$) with copula C. The multivariate copula C is given by

$$C(u_1, u_2 \ \ldots, u_p) = \int_{-\infty}^{t_v^{-1}(u_1)} \ldots \int_{-\infty}^{t_v^{-1}(u_p)} f(t)dt$$

t_v^{-1} is the inverse of the marginal distribution of the marginals, $f(t)$ denotes the copula function (i.e. for a t-copula it's a student's t distribution) [144, 145].

The standard formulation of a t-copula with two continuous random variables X_1, X_2 is defined as follows:

$$\boldsymbol{C(F_1(X_1), F_2(X_2))} \;=\; \int_{-\infty}^{t_v^{-1}(F_1(X_1))} \int_{-\infty}^{t_v^{-1}(F_1(X_2))} \frac{\mathbf{1}}{\mathbf{2\pi(1-\rho^2)^{1/2}}} \Big\{ \mathbf{1} + \frac{\mathbf{x_1{}^2 - 2\rho x_1 x_2 + x_2{}^2}}{\boldsymbol{\nu(1-\rho^2)}} \Big\}^{-(\nu+2)/2} \boldsymbol{dt} \qquad \textbf{(2.1)}$$

where C is the copula, and F_1 and F_2 are marginal distribution functions, ρ and ν are the parameters of the copula to be set during training, x_1, x_2 are samples sampled from the distributions $F_1(X_1), F_2(X_2)$, and t_v^{-1} is the inverse of the standard univariate student-t-distribution with v degrees of freedom, expectation 0 and variance $\frac{v}{v-2}$ [145, 146].

The continuous copula distribution can be converted into discrete copula using the methods described in the paper [146]. In our formulations, we used MATLAB implementation of copulafit to fit the copula and sample from the copula [146].

Hence, each feature with missing data Y_i is then sampled from a distribution given by

$$Y_i \sim C(F_1(X_1), F_2(X_2), \dots, F_{\mathrm{N}}(X_N), F_{\mathrm{N+i}}(R_i)) \qquad (2.2)$$

where $X = [X_1, X_2, \dots, X_N]$ is the data with N features and R_i is the missingness pattern for feature Y_i. The parameters for the copula are maximum likelihood estimates fit using observed data and the "missingness" pattern at a p-value of 0.05, 0.10 and 0.15.

Evaluation of the Imputation Methods using Random Forests

We evaluate all the heretofore mentioned methods for non-temporal analysis using Random Forests to predict ICU mortality on the three datasets and sepsis on two datasets (since the number of sepsis patients for the pediatric data was very small). The new

methods for imputations were tested against conventional expectation maximization, mean filling and no filling, using Random Forests to predict mortality in the ICU. Random Forests was chosen due to its robustness to missing data. The scores used for evaluation were accuracy and Mathews correlation coefficient (MCC) and 3 × 3 nested cross-validation [147]. MCC was chosen as an evaluation since it's relatively insensitive to an imbalance in the population

Feature Interpretation

For each imputation technique, we report the best performing features in terms of their importance scores [148]. We also report the number of features which contributed to 90% of the prediction for each of the imputation techniques. The importance scores for each feature m from a total of M are computed as follows (Figure. 2.2):

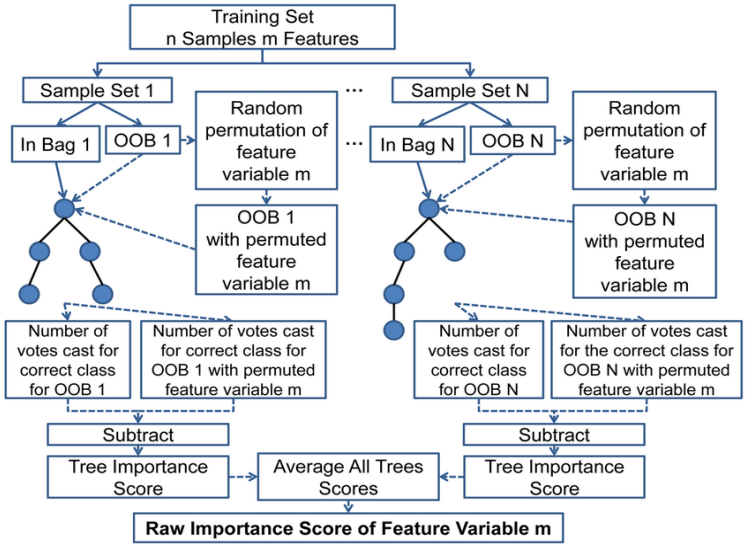


Figure 2.2: Feature Interpretation: Calculating Importance Scores

- For a total of N trees, N samples of the training data is generated.
- For each tree, the data is divided into in the bag samples (used for tree generation) and the our of bag samples (OOB) (used for testing the tree).

- The OOB samples for the feature m are permuted and the increase in error is computed as the tree importance score.
- This value is averaged across all the trees to obtain the raw importance score used to rank all the features.

2.4. Results

We demonstrate our results for retrospective data analysis using data from both Multi-parameter Intelligent Monitoring in Intensive Care, second version, (MIMIC-II) database [149] and 'Medical Information Mart for Intensive Care database (MIMIC-III)[150]

2.4.1 Case Study 1a: Adult ICU Database – MIMIC-II

Data Source – MIMIC-II Database

MIMIC-II is a public ICU data repository containing over 40,000 ICU stay records (32,331 adult and 8080 neonatal records) [149]. The MIMIC II data for each patient is either static (does not change over the entire duration of the patient ICU stay, e.g., patient demographics) or temporal (changing in time, e.g., heart rate, blood pressure). From a total 13,000 features in MIMIC-II database, we ranked the features by the number of available records. From the top 2000 features, we picked 87 features with the greatest clinical significance (based on clinician input). These included measures of physiological parameters (e.g. heart rate, blood pressure), lab results (e.g. WBC, RBC, cholesterol), administrative data (e.g. length of stay, ICD-9), comorbidities and other diagnostic procedures. On this dataset, we performed two type analysis, temporal and non-temporal.

Non-Temporal Data Analysis

For non-temporal analysis, we used the temporal data by converting it into values averaged over the duration of stay. Then outliers whose values were physiologically impossible were removed. If the value is normally distributed, then values that deviated by ±3 standard deviations from the mean value were also removed. After preprocessing and outlier removal, the total missing data in the dataset was 30.05% (mean) with a standard deviation of 30.8. We performed our analysis using adult data from the MIMIC II database, which consists of 32,331 adult records. In this dataset, there were 2,334 patient records with mortality during the ICU stay and 29,997 patient records of successful discharge from the ICU. The missing data was 30.6% (mean) and 30.9 (standard deviation) in patients with successful discharge from the ICU and 22.47 (mean) and 30.44 (standard deviation) in patients with ICU mortality. In this dataset, there were 2,010 patient records with sepsis during the ICU stay and 29,947 patient records with no sepsis in the ICU. The missing data was 25.82% (mean) and 26.39 (standard deviation) in patients with no sepsis in the ICU and 29.58 (mean) and 31.61 (standard deviation) in patients with ICU sepsis.

Temporal Data Analysis

For non-temporal analysis, we used the temporal data by converting it into values averaged over the duration of stay and for temporal analysis, the data was binned into with sampling intervals of 2, 6 and 12 hours. If the data was missing then no substitutions were made until aim1. Then outliers whose values were physiologically impossible were removed. If the value is distributed normally, then values which deviated by ±3 standard deviations from the mean value were removed. After preprocessing and outlier removal, the total missing data in the dataset was about 87% (mean) and standard deviation 21. We first distinguish the missing data into sampling related and true missing. A data is called

true missing if the sampling interval between any two points is greater than 2 times the 95th percentile of all sampling interval, else it is sampling related. All sampling related missing data is imputed using 6 different techniques (cubic, linear, and piecewise linear, expectation maximization, nearest neighbor and spline interpolation) and the true missing data is handled as mentioned above.

Test for "Neglectable" Assumption

We performed the t-tests and Little's test. The results of the t-test from non-temporal analysis (Figure 2.3) demonstrate that most of the p-values reject the null hypothesis which states that the means of the two sample populations are derived from same distribution, indicating the data is not "Neglectable." It is supported by the results of the Little's test which showed that the dataset is not "Neglectable" (batch size =5 in Table 2.2).

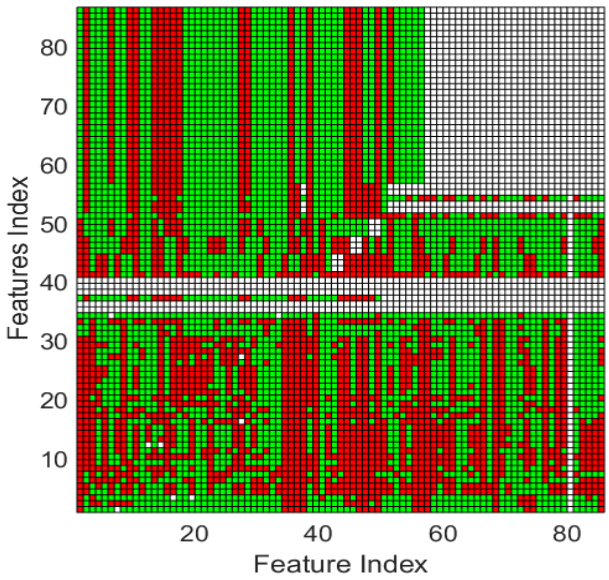


Figure 2.3: t-test results: Green row means that particular feature is "Neglectable". Since Green row means that particular feature is "Neglectable". No color means no missing data in feature or the missing data pattern is same as test feature. Since no row is green, data is not "Neglectable"

Table 2.2: Little's test results (Batch-Size = 5)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-5	6023.27	40	0
Feat 6-10	2062.79	75	0
Feat 11-15	1382.4	55	0
Feat 16-20	2275.72	32	0
Feat 21-25	1062.145	73	0
Feat 26-30	1767.2	53	0
Feat 31-35	409.22	46	0
Feat 36-40	2802.38	27	0
Feat 41-45	3020.52	4	0
Feat 46-50	3376.06	15	0
Feat 51-55	993.54	5	0
Feat 56-60	1617.11	11	0
Feat 61-65	7.134	1	0.007
Feat 66-87	12.549	3	0.005

For temporal analysis, since the records of each time point are not independent, and t test assumes independence, we performed the t-test against the means for each patient The temporal analysis with 6 different interpolation techniques for imputing sampling related missing data, also proved that data is not "Neglectable". Figure 2.4 shows the results

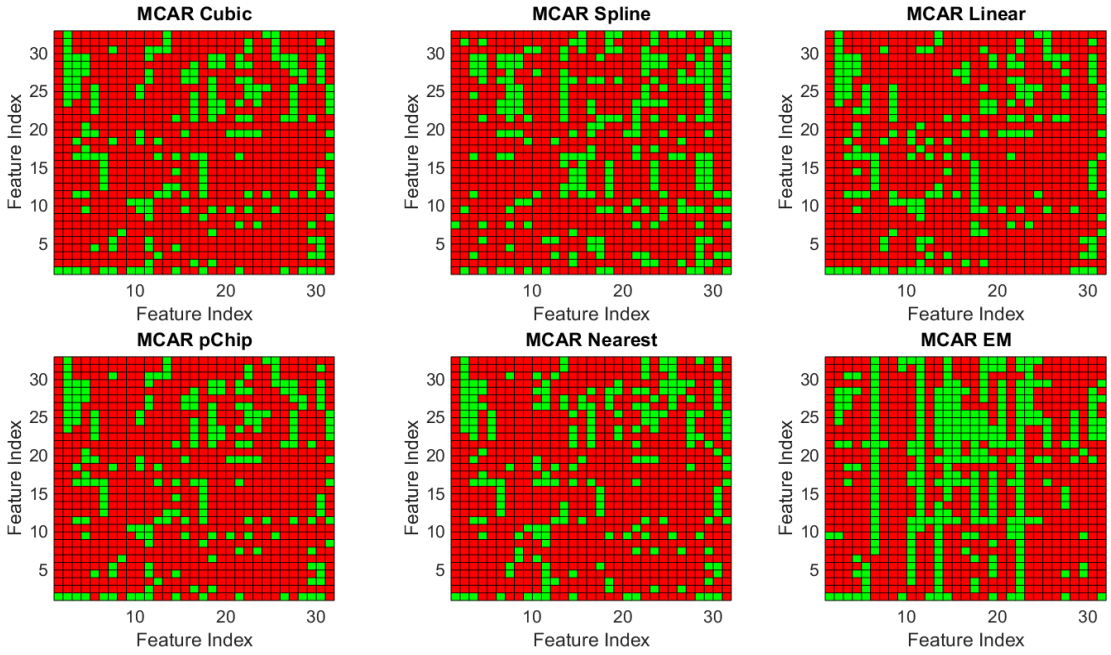


Figure 2.4: t-test results - Temporal: Green row means that particular feature is "Neglectable". These results suggest data is not "Neglectable" a) Cubic b) Spline c) Nearest d) Linear e) Piecewise cubic f) Expectation maximization.

Table 2.3: Little's test results indicate data is not "Neglectable"

	Cubic			Spline			Nearest			Linear			pChip			EM		
Feature #	**Chi Square**	**DF**	**pVal**	**Chi Square**	**DF**	**pVal**	**Chi Square**	**DF**	**pVal**	**Chi Square**	**DF**	**pVal**	**Chi Square**	**DF**	**pVal**	**Chi Square**	**DF**	**pVal**
Feat 1-5	679.98	27	0	1130.78	56	0	1363.07	60	0	1395.91	59	0	1421.33	59	0	1575.81	59	0
Feat 6-10	419.30	28	0	869.35	75	0	1805.19	75	0	2361.28	75	0	2028.26	75	0	1615.81	75	0
Feat 11-15	518.14	27	0	787.57	51	0	815.62	54	0	924.42	58	0	1660.13	59	0	317.53	58	0
Feat 16-20	1329.4	28	0	1167.47	71	0	1759.98	73	0	2061.83	73	0	2014.74	73	0	2335.47	73	0
Feat 21-25	326.39	28	0	413.56	64	0	217.72	58	0	452.57	61	0	803.47	61	0	371.88	61	0
Feat 26-30	163.36	28	0	174.91	75	0	417.13	71	0	528.68	75	0	491.58	75	0	1146.98	75	0
Feat 31-33	1447.29	75	0	24.85	9	0	126.86	9	0	209.89	9	0	206.08	9	0	181.35	9	0

of 6 of the interpolation techniques for 6hr. The results for 2hr. and 12hr. intervals were also similar. This is supported by the results from the Little's test also where all groups showed that the results were not "Neglectable" (Table 2.3).

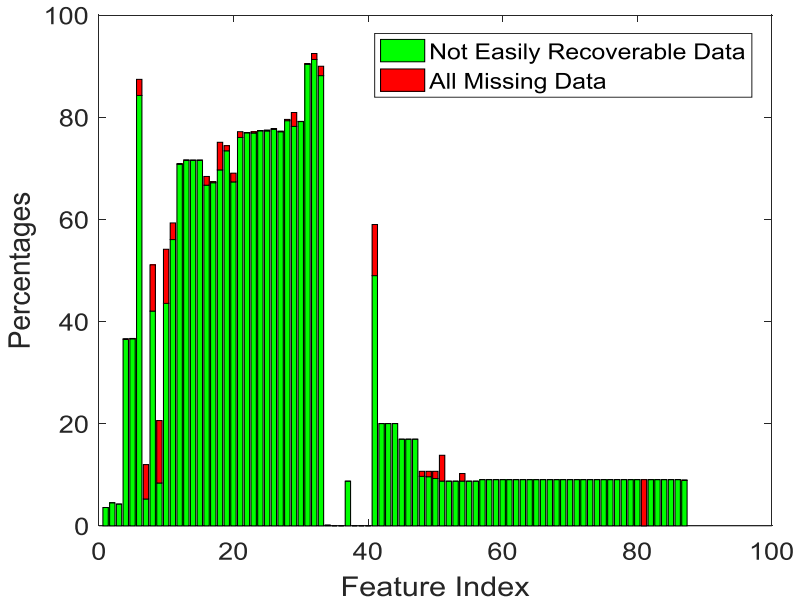


Figure 2.5: "NER" data identification: The red bar gives percentage of all missing data the green bars give the "NER" data percentage. The features which have no missing data do not have any bars.

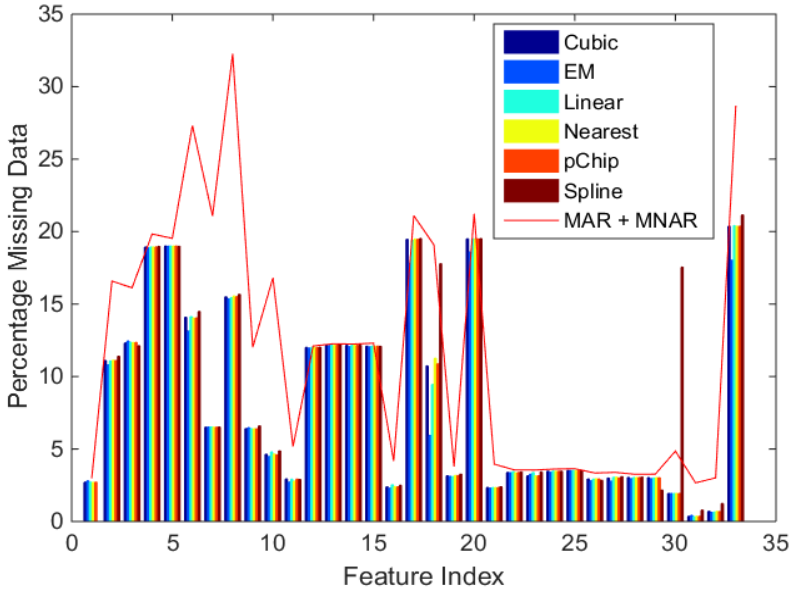


Figure 2.6: MNAR Data Identification - Temporal: The red line gives the percentage of true missing data and the bars give the MNAR data percentage for each of the 6 imputation methods.

Identifying "NER" Data

The classification analysis (Figure 2.5) shows very high levels of the missing data to be "NER."(33.2% of the data and 99% of all missing data) for non-temporal analysis. Figure 2.6 shows the results of temporal "NER" identification, it shows that though "NER" is lower than non-temporal data, it still constitutes 70% of true missing data. Also the results are similar across binning intervals. These results indicate that most conventional approaches of imputing all the data using "Recoverable" assumptions or deleting may lead to bias. In temporal analysis, the percentages of "NER" were found to be lower due to a higher correlation between adjacent time points. However, the high levels of "NER" make the prediction and data interpretation challenging. Therefore, we perform estimation of data under "NER" assumptions.

Evaluation using Random Forests

Evaluation of the imputation models was performed on the non-temporal data using Random Forests to predict ICU mortality and sepsis. The list of all features along with the category is found in the Appendix.

Results for Mortality Prediction

The models where "NER" data was imputed using copulas outperformed all the other models. The k-means based "Recoverable" models outperformed traditional EM models, mean filling and no filling techniques (Figures 2.7, 2.8). The statistical significance of these models was tested using Steigler's Z score [151] for correlated correlations from the MCC scores.

On comparing the prediction performance of our novel methods for statistical significance, we found that all the novel models which impute "NER" outperformed EM

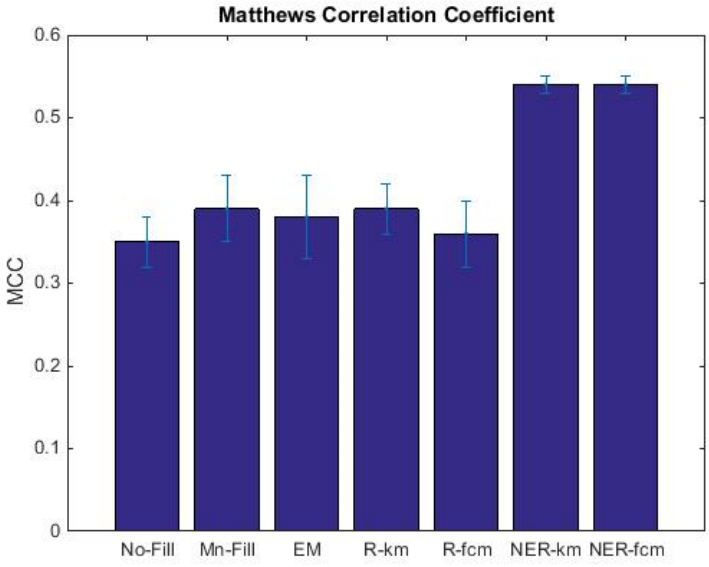


Figure 2.7: Mortality prediction results- MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). NER models gave best performance.

algorithm and mean filling imputation techniques with a statistical significance of $p \leq .01$. All the proposed novel data handling techniques were shown to be performing better than

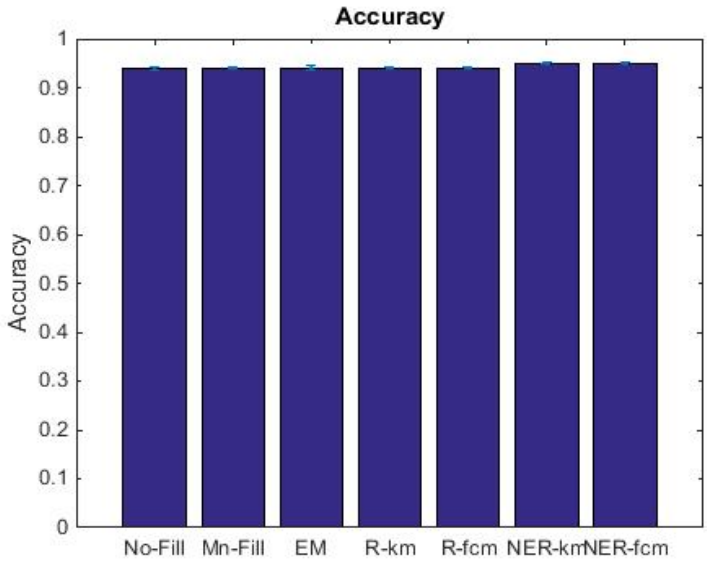


Figure 2.8: Mortality prediction results - Accuracy (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). NER models gave best performance.

no data imputation with a statistical significance of $p \leq .01$. The repetitions of NER methods with the different p-value parameters (0.05, 0.10 and 0.15) all gave MCC values greater than 0.55 and accuracy greater than 0.95.

Table 2.4: Top 10 MIMIC-II mortality features in each of the models with the importance scores and the number of features accounting for 90% of the total importance. All the features in the order of importance is given in the appendix. The features in bold have also been reported by other studies using MIMIC-II to be indicative of mortality.

SN	No Fill	Imp 90% features = 56	Mn Fill	Imp 90% features = 58	EM	Imp 90% features = 57	kmeans Rec	Imp 90% features = 62	fcm Rec	Imp90% features = 58	kmeans NER	Imp 90% features = 56	fcm NER	Imp90% features = 59
1	**Simplified Acute Physiology Score I Min**	0.0341	Sequential Organ Failure Assessment Min	0.0387	**Simplified Acute Physiology Score I Max**	0.0388	**Simplified Acute Physiology Score I Min**	0.0389	**Simplified Acute Physiology Score I Max**	0.0406	Sequential Organ Failure Assessment Min	0.0483	Hospital Length of Stay	0.0439
2	**SpO2**	0.0341	**Simplified Acute Physiology Score I Max**	0.0387	Sequential Organ Failure Assessment Min	0.0388	**SpO2**	0.0389	hospital Length of Stay	0.0406	Hospital Length of Stay	0.0483	**Simplified Acute Physiology Score I Max**	0.0439
3	Hospital Length of Stay	0.0341	Sequential Organ Failure Assessment Max	0.0387	Sequential Organ Failure Assessment Max	0.0388	Sequential Organ Failure Assessment Min	0.0389	Sequential Organ Failure Assessment min	0.0406	**SpO2**	0.0483	Sequential Organ Failure Assessment Max	0.0439
4	Icustay Length of Stay	0.0341	**SpO2**	0.0387	Hospital Length of Stay	0.0388	Sequential Organ Failure Assessment Max	0.0389	**Simplified Acute Physiology Score I Min**	0.0406	Icustay Length of Stay	0.0483	Sequential Organ Failure Assessment Min	0.0439
5	Cost Weight	0.0341	**Simplified Acute Physiology Score I Min**	0.0387	**Simplified Acute Physiology Score I Min**	0.0388	Simplified Acute Physiology Score I Max	0.0389	icustay Length of Stay	0.0406	Sequential Organ Failure Assessment Max	0.0483	**SpO2**	0.0439
6	**Arterial BP**	0.0341	**Arterial BP**	0.0387	Cost Weight	0.0388	Icustay Length of Stay	0.0389	Sequential Organ Failure Assessment max	0.0406	**Simplified Acute Physiology Score I Min**	0.0483	**Simplified Acute Physiology Score I Min**	0.0439
7	Sequential Organ Failure Assessment Min	0.0341	Hospital Length of Stay	0.0387	Sequential Organ Failure Assessment First	0.0388	Hospital Length of Stay	0.0389	**SpO2**	0.0406	Simplified Acute Physiology Score I Max	0.0483	Cost Weight	0.0439
8	Sequential Organ Failure Assessment Max	0.0341	Cost Weight	0.0387	Icustay Length of Stay	0.0388	Cost Weight	0.0389	**Heart Rate**	0.0406	Simplified Acute Physiology Score I First	0.0483	Icustay Length of Stay	0.0439
9	Simplified Acute Physiology Score I Max	0.0341	**Heart Rate**	0.0387	Subject Total Number of ICU Stays	0.0388	Simplified Acute Physiology Score I First	0.0389	Simplified Acute Physiology Score I first	0.0406	Cost Weight	0.0483	Sequential Organ Failure Assessment First	0.0439
10	Simplified Acute Physiology Score I First	0.0341	Glucose (70-105)	0.0387	Simplified Acute Physiology Score I First	0.0388	**Arterial BP Mean**	0.0389	**Arterial BP Mean**	0.0406	**Heart Rate**	0.0483	**Heart Rate**	0.0439

These results prove that division of missing data into "Neglectable", "Recoverable" and "NER" and the novel imputation methods give a better performance as compared to current strategies of EM, mean filling, and no filling.

Interestingly, the features that were seen to be most indicative of mortality are very similar irrespective of the imputation methods. Top ranking features predicted using our model (Table 2.4) such as SAP scores, long length of ICU stay, SpO_2, comorbidities and SOFA scores have been clinically shown to be correlated with mortality [149, 152-158]. The features such as SAPS-I [149], ABP [159], age, heart rate, systolic blood pressure, body temperature, Glasgow Coma Scale, mechanical ventilation, PaO2, FiO2, urine output, BUN (blood urea nitrogen), blood sodium, potassium, bicarbonates, bilirubin, white blood cells, chronic disease (AIDS, metastatic cancer, hematologic malignancy) and type of admission (elective surgery, medical, unscheduled surgery)[160] have been shown to be associated with mortality from other studies using the MIMIC-II dataset. In addition, SAPS-I, SpO_2, creatinine have been shown to associated with mortality in sepsis patients [157].

Results for Sepsis Prediction

The models where "NER" data was imputed using copulas outperformed all the other models except mean filling. The "Recoverable" models outperformed traditional EM models, and no filling techniques (Figures 2.9, 2.10). The statistical significance of these models was tested using Steigler's Z score [151] for correlated correlations from the MCC scores and the results were not statistically significant at $p \leq .01$.

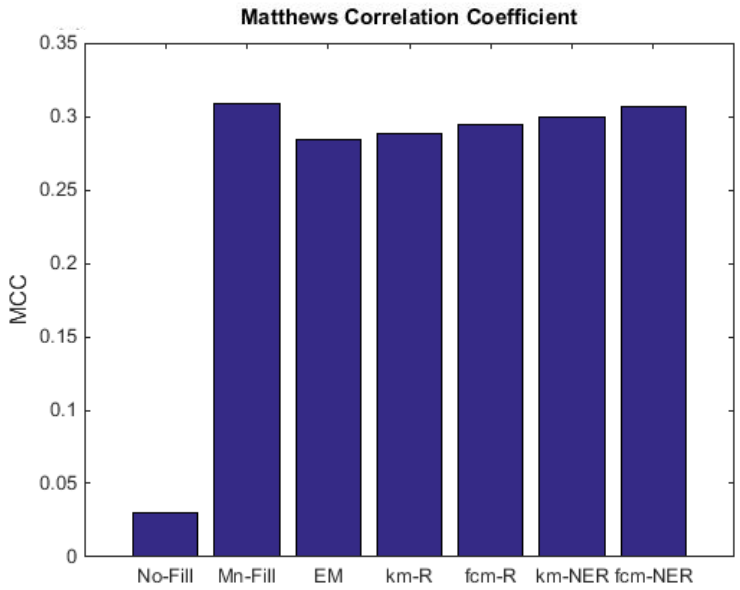


Figure 2.9: Sepsis prediction results- MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). Excluding mean filling, NER models gave best performance.

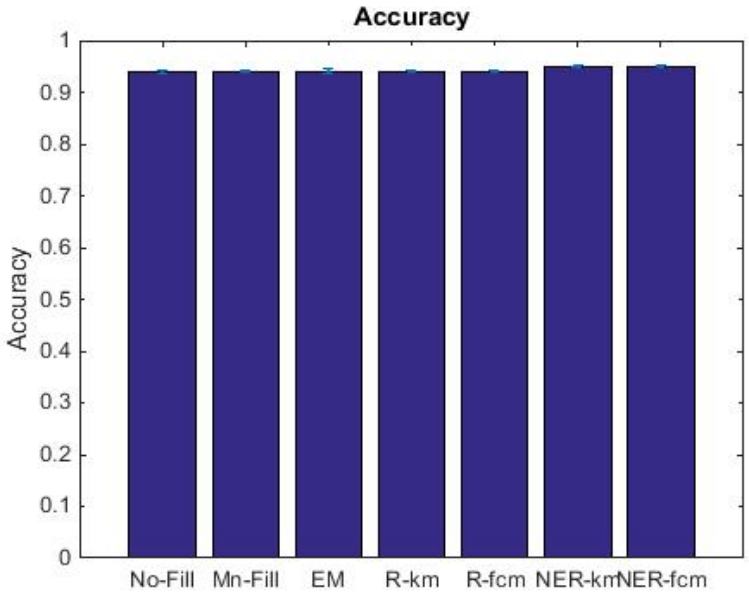


Figure 2.10: Sepsis prediction results- Accuracy (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). Excluding mean filling, NER models gave best performance.

These results prove that division of missing data into "Neglectable", "Recoverable"

and “NER” and the novel imputation methods give a better performance as compared to current strategies of EM, and no filling for sepsis.

Top ranking features predicted using our model (Table 2.5) such as SOFA scores , long length of ICU stay, blood pressure, pulse pressure, heart rate, temperature, respiration rate, white blood cell count, pH, blood oxygen saturation and age have been clinically shown to be correlated with sepsis [161-164]. The features such as blood pressure, pulse pressure, heart rate, temperature, respiration rate, white blood cell count, pH, blood oxygen saturation, and age have been clinically shown to be correlated with sepsis from other studies using the MIMIC-II dataset [164].

The best performing models from non-temporal analysis models (NER with kmeans and fcm) were used for temporal classification and evaluation of temporal models was performed using conditional random fields. The results are detailed in chapter 2 and are compared against logistic regression and feed-forward neural networks used as baselines.

Table 2.5: Top 10 MIMIC-II sepsis features in each of the models with the importance scores and the number of features accounting for 90% of the total importance. All the features in the order of importance is given in the appendix. The features in bold have also been reported by other studies using MIMIC-II to be indicative of sepsis.

SN	**No Fill**	**Imp 90 % features = 62**	**Mn Fill**	**Imp 90 % features = 65**	**EM**	**Imp 90 % features = 59**	**kmeans Rec**	**Imp 90 % features = 62**	**fcm Rec**	**Imp 90 % features = 59**	**kmeans NER**	**Imp 90 % features = 62**	**fcm NER**	**Imp 90 % features =67**
1	Height	0.0317	Cost Weight	0.0352	Cost Weight	0.0285	Height	0.0303	ICU Stay Length of Stay	0.0268	**Sequential Organ Failure Assessment First**	0.0361	**Sequential Organ Failure Assessment Max**	0.0283
2	**SaO2**	0.0273	**Sequential Organ Failure Assessment Max**	0.0233	**ICU Stay Admit Age**	0.0265	**Sequential Organ Failure Assessment First**	0.0268	**SaO2**	0.0263	**Sequential Organ Failure Assessment Max**	0.0318	Hospital Length of Stay	0.0255
3	**Sequential Organ Failure Assessment Max**	0.0257	CVP	0.0232	Lactic Acid	0.0231	Glucose (70-105)	0.0251	Calcium (8.4-10.2)	0.0253	Cost Weight	0.0282	Cost Weight	0.0255
4	**Heart Rate**	0.0241	**ICU Stay Admit Age**	0.0230	Weight Min	0.0221	**Sequential Organ Failure Assessment Max**	0.0247	Height	0.0242	Hospital Length of Stay	0.0269	**Sequential Organ Failure Assessment First**	0.0248
5	Glucose (70-105)	0.0230	**NBP**	0.0230	CVP	0.0218	Arterial BP	0.0246	Creatinine (0-1.3)	0.0241	**Heart Rate**	0.0238	**Heart Rate**	0.0246
6	Simplified Acute Physiology Score I Max	0.0228	**Heart Rate**	0.0225	**NBP**	0.0218	Carbon Dioxide	0.0230	Fingerstick Glucose	0.0233	**NBP**	0.0232	**ICU Stay Admit Age**	0.0207
7	ICU Stay Length of Stay	0.0218	Total Number of ICU Stays	0.0217	Hospital Length of Stay	0.0217	Cost Weight	0.0220	CVP	0.0228	**Respiratory Rate**	0.0214	Fluid Electrolyte Amount	0.0207
8	Cost Weight	0.0218	Hospital Length of Stay	0.0203	**SpO2**	0.0216	Hospital Length of Stay	0.0217	Religion	0.0210	Total Number of Hospital Stays	0.0209	Simplified Acute Physiology Score I First	0.0207
9	**ICU Stay Admit Age**	0.0218	**Respiratory Rate**	0.0197	**WBC**	0.0213	Simplified Acute Physiology Score I Max	0.0213	**Hematocrit**	0.0208	Fluid Electrolyte Amount	0.0197	Lymphoma (Y/N)	0.0202
10	Hospital Length of Stay	0.0211	ICU Stay Length of Stay	0.0196	Sodium (135-148)	0.0192	Arterial CO2(Calc)	0.0207	**Heart Rate**	0.0207	Simplified Acute Physiology Score I Max	0.0196	ICU Stay Length of Stay	0.0200

2.4.2 Case Study 1b: Adult ICU Database – MIMIC-III

Data Source – MIMIC-III Database

MIMIC-III is a public ICU data repository containing 52,963 adult ICU stay records. The MIMIC III data for each patient is either static (does not change over the entire duration of the patient ICU stay, e.g., patient demographics) or temporal (changing in time, e.g., heart rate, blood pressure). From a total 13,000 features in MIMIC-III database, we ranked the features by the number of available records. From the top 2000 features, we picked 147 features with the greatest clinical significance (based on clinician input) and frequency. These included measures of physiological parameters (e.g. heart rate, blood pressure), lab results (e.g. WBC, RBC, cholesterol), administrative data (e.g. length of stay, ICD-9), comorbidities and other diagnostic procedures. On this dataset, we performed non-temporal analysis only.

Non-Temporal Data Analysis

For non-temporal analysis, we used the temporal data by converting it into values averaged over the duration of stay. Then outliers whose values were physiologically impossible were removed. If the value is normally distributed, then values that deviated by ±3 standard deviations from the mean value were also removed. After preprocessing and outlier removal, the total missing data in the dataset was 18.57% (mean) with a standard deviation of 26. In this dataset, there were 4,726 patient records with sepsis during the ICU stay and 48,237 patient records with no sepsis in the ICU. The missing data was 17.6% (mean) and 25.65 (standard deviation) in patients with no sepsis in the ICU and 18.67 (mean) and 26.57 (standard deviation) in patients with ICU sepsis. In this dataset, there

were 6,531 patient records with mortality during the ICU stay and 46,432 patient records with no mortality in the ICU. The missing data was 17.69% (mean) and 24.50 (standard deviation) in patients with no mortality in the ICU and 18.70 (mean) and 26.80 (standard deviation) in patients with ICU mortality.

Test for "Neglectable" Assumption

We performed the t-tests and Little's test. The results of the t-test from non-temporal analysis (Figure 2.11) demonstrate that most of the p-values reject the null hypothesis which states that the means of the two sample populations are derived from same distribution, indicating the data is not "Neglectable." It is supported by the results of the Little's test which showed that the dataset is not "Neglectable" (Chi-Square = 936940.041, DF = 667324, Sig. = .000).

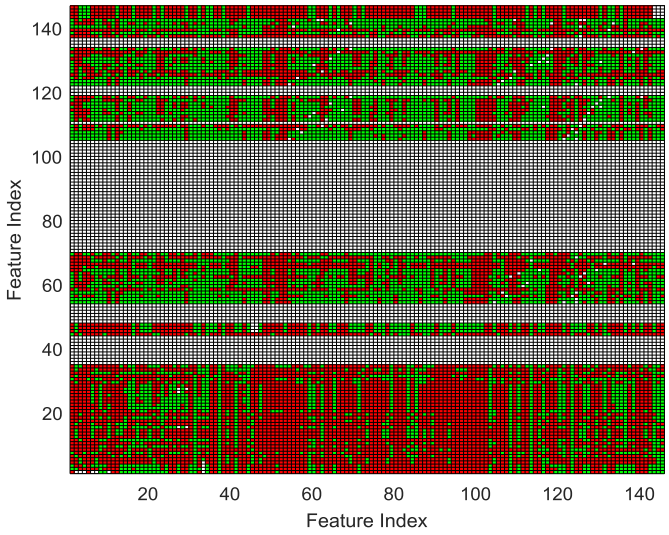


Figure 2.11: t-test results: Green row means that particular feature is "Neglectable". Since Green row means that particular feature is "Neglectable". No color means no missing data in feature or the missing data pattern is same as test feature. Since no row is green, data is not "Neglectable"

Identifying “NER” Data

The classification analysis (Figure 2.12) shows very high levels of the missing data to be “NER.”(17% of the data and 95% of all missing data). These results indicate that most conventional approaches of imputing all the data using “Recoverable” assumptions or deleting may lead to bias. to be lower due to a higher correlation between adjacent time points. However, the high levels of “NER” make the prediction and data interpretation challenging. Therefore, we perform estimation of data under “NER” assumptions.

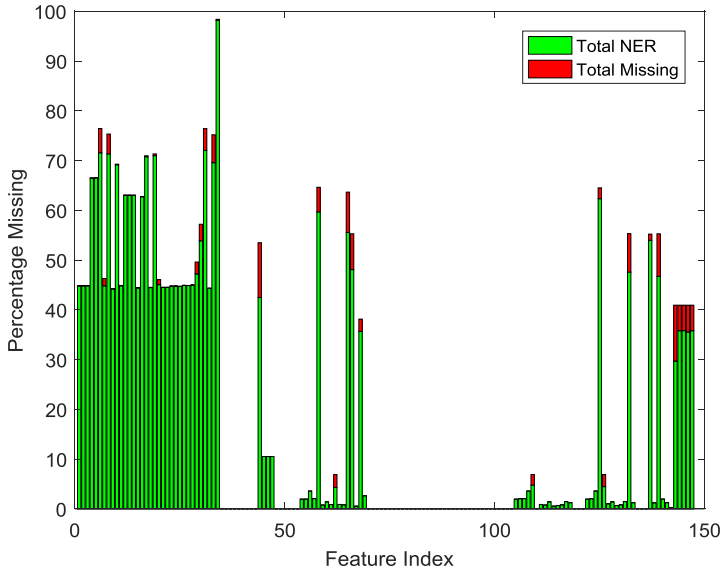


Figure 2.12: “NER” data identification: The red bar gives percentage of all missing data the green bars give the “NER” data percentage. The features which have no missing data do not have any bars.

Evaluation using Random Forests

Evaluation of the imputation models was performed on the non-temporal data using Random Forests to predict ICU mortality and sepsis. The list of all features along with the category is found in the Appendix.

Results for Mortality Prediction

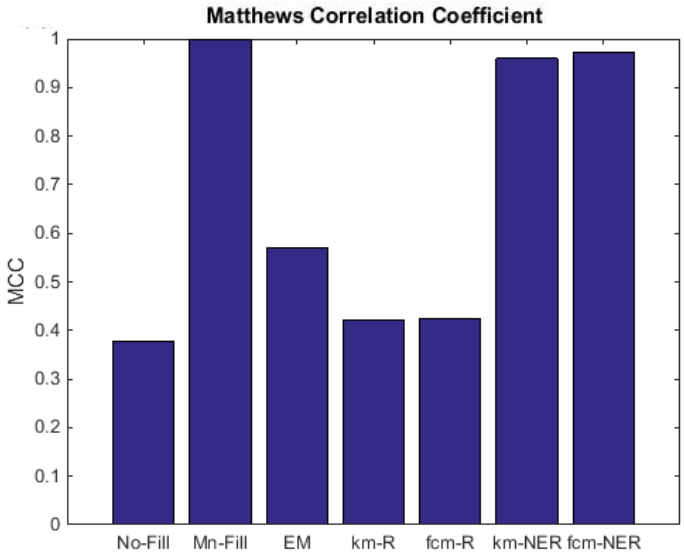


Figure 2.13: Mortality prediction results- MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). Excluding mean filling, NER models gave best performance.

The models where "NER" data was imputed using copulas outperformed all the other models except mean filling (Figures 2.13, 2.14). The statistical significance of these

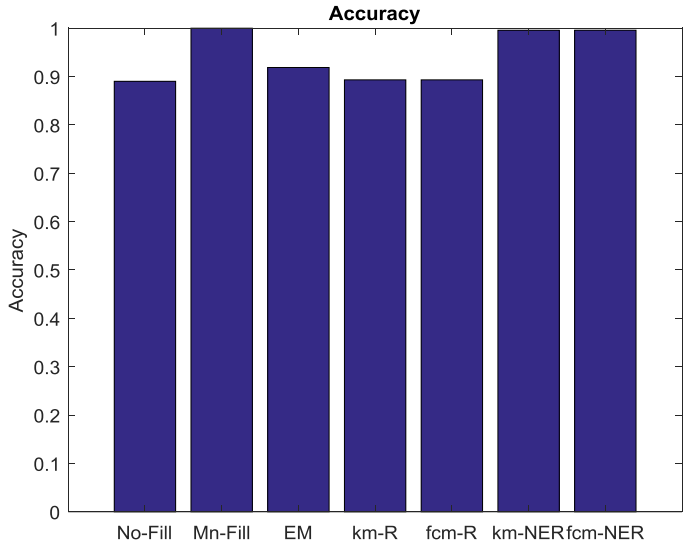


Figure 2.14: Mortality prediction results- MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). Excluding mean filling, NER models gave best performance.

Table 2.6: Top 10 MIMIC-III mortality features in each of the models with the importance scores and the number of features accounting for 90% of the total importance. All the features in the order of importance is given in the appendix. The features in bold have also been reported by other studies using MIMIC-III to be indicative of mortality.

SN	No Fill	Imp 90% features = 112	Mn Fill	Imp 90% features = 112	EM	Imp 90% features = 105	kmeans Rec	Imp 90% features = 108	fcm Rec	Imp 90% features = 104	kmeans NER	Imp 90% features = 104	fcm NER	Imp 90% features =101
1	Discharge Location	0.032154	Discharge Location	0.077526	**SpO2**	0.019813	Discharge Location	0.026501	Discharge Location	0.031866	Discharge Location	0.063089061	Discharge Location	0.062203
2	Respiration Rate (Total)	0.014491	**SpO2**	0.016312	Lactic Acid	0.019104	**Arterial BP [Systolic]**	0.018332	**Arterial BP [Systolic]**	0.015401	**Arterial BP [Systolic]**	0.021923849	**Arterial BP [Systolic]**	0.019909
3	Simplified Acute Physiology Score II Heart Rate Score	0.014443	Lactic Acid	0.013935	**Arterial BP [Systolic]**	0.016555	**Arterial BP Mean**	0.017017	**Arterial BP Mean**	0.014109	**Arterial BP Mean**	0.018364994	**Arterial BP Mean**	0.015866
4	Weight Min	0.013646	**Age**	0.013561	**Mechanical Ventilationilation (Y/N)**	0.014777	Sequential Organ Failure Assessment Liver	0.015405	Sequential Organ Failure Assessment Liver	0.013888	Sequential Organ Failure Assessment Liver	0.017311465	Sequential Organ Failure Assessment Liver	0.015747
5	Sequential Organ Failure Assessment Liver	0.013431	Elixhauser Ahrq Score Elixhauser Vanwalraven	0.012472	**Arterial pH**	0.014635	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	0.014991	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	0.013707	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	0.015677133	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	0.015698
6	Arterial PaCO2	0.013286	Mechanical Ventilationilation (Y/N)	0.012116	Hemoglobin	0.014304	Simplified Acute Physiology Score II	0.013997	Simplified Acute Physiology Score II	0.012733	Simplified Acute Physiology Score II	0.015201917	Simplified Acute Physiology Score II	0.015186
7	Lactic Acid	0.013072	Simplified Acute Physiology Score	0.011699	**Sodium (135-148)**	0.014118	Respiration Rate (Total)	0.013795	Respiration Rate (Total)	0.012502	Respiration Rate (Total)	0.014811586	Respiration Rate (Total)	0.015045
8	**Potassium (3.5-5.3)**	0.013069	Request Respiratory Monitoring Periodically (Y/N)	0.011362	NBP [Systolic]	0.013654	Albumin (>3.2)	0.013287	Albumin (>3.2)	0.012512	Albumin (>3.2)	0.013485976	Albumin (>3.2)	0.014696
9	**Arterial BP Mean**	0.012168	Simplified Acute Physiology Score II Prob	0.011266	Carbon Dioxide	0.013508	Acute Physiology Score III Bilirubin Score	0.013214	Acute Physiology Score III Bilirubin Score	0.012335	Acute Physiology Score III Bilirubin Score	0.013414385	Acute Physiology Score III Bilirubin Score	0.013776
10	SaO2	0.012066	Carbon Dioxide	0.010564	**WBC (4-11,000)**	0.013305	Weight First	0.013153	Weight First	0.012222	Weight First	0.012608929	Weight First	0.013652

models was tested using Steigler's Z score [151] for correlated correlations from the MCC scores.

On comparing the prediction performance of our novel methods for statistical significance, we found that all the novel models which impute "NER" outperformed EM algorithm and no filling imputation techniques with a statistical significance of $p \leq .01$. These results prove that division of missing data into "Neglectable", "Recoverable" and "NER" and the novel imputation methods give a better performance as compared to current strategies of EM, mean filling, and no filling.

Top ranking features predicted using our model (Table 2.6) such as SAP scores, long length of ICU stay, SpO_2, comorbidities and SOFA scores have been clinically shown to be correlated with mortality [149, 152-158]. The features such as Glasgow Coma Scale, systolic blood pressure, diastolic blood pressure, heart rate, temperature, respiration rate, SpO_2, urine output, FiO_2, blood pH, total bilirubin, creatinine, platelets, white blood cell count, serum bicarbonate, sodium, potassium, age, $PaO_2$and comorbidities [165-167] have been shown to be associated with mortality from other studies using the MIMIC-III dataset.

Results for Sepsis Prediction

The models where "NER" data was imputed using copulas outperformed all the other models except mean filling. The "Recoverable" models outperformed traditional EM models, and no filling techniques (Figures 2.15, 2.16). The statistical significance of these

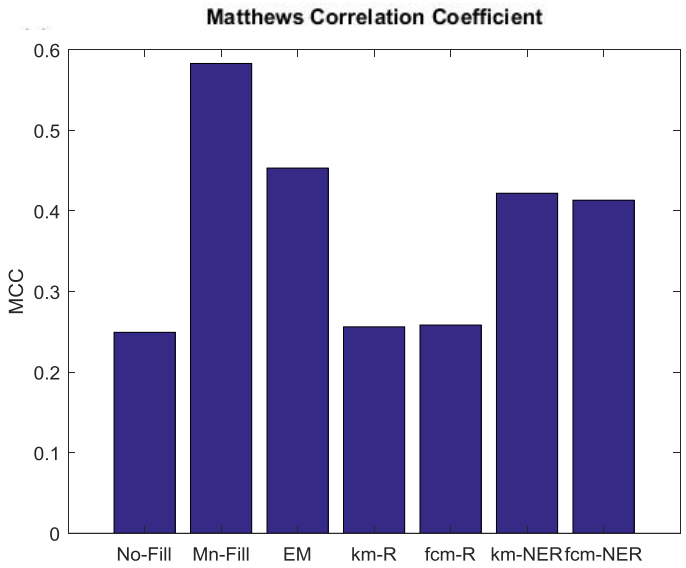


Figure 2.15: Sepsis prediction results- MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation

models was tested using Steigler's Z score [151] for correlated correlations from the MCC scores and the results were not statistically significant at $p \leq .01$.

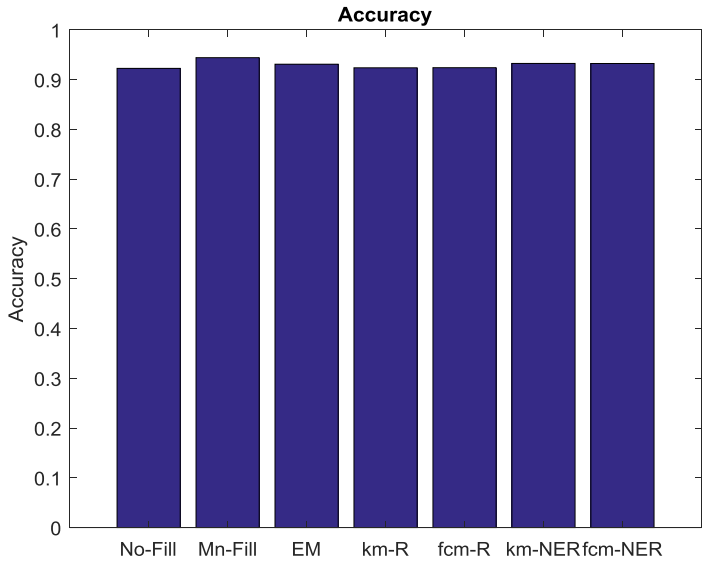


Figure 2.16: Sepsis prediction results- MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation).

These results prove that division of missing data into “Neglectable”, “Recoverable” and “NER” and the novel imputation methods give a better performance as compared to current strategies of no filling for sepsis.

Top ranking features predicted using our model (Table 2.7) such as SOFA scores , long length of ICU stay, blood pressure, pulse pressure, heart rate, temperature, respiration rate, white blood cell count, pH, blood oxygen saturation and age have been clinically shown to be correlated with sepsis [161-164]. The features such as Glasgow Coma Scale, heart rate, respiration rate, SpO2, blood pressure, and temperature have been clinically shown to be correlated with sepsis from other studies using the MIMIC-III dataset [168].

Table 2.7: Top 10 MIMIC-III Sepsis Features in each of the Models with the Importance Scores and the Number of Features accounting for 90% of the Total Importance. All the Features in the Order of Importance is given in the Appendix. All the features in the order of importance is given in the appendix. The features in bold have also been reported by other studies using MIMIC-III to be indicative of mortality.

SN	No Fill	Imp 90% features = 110	Mn Fill	Imp 90% features = 115	EM	Imp 90% features = 109	kmeans Rec	Imp 90% features =111	fcm Rec	Imp 90% features = 112	kmeans NER	Imp 90% features = 113	fcm NER	Imp 90% features = 108
1	CVP	0.014221	Sequential Organ Failure Assessment Cardiovascular	0.021235	Sequential Organ Failure Assessment Cardiovascular	0.016919	SaO2	0.014982	Lactic Acid	0.015485	Sequential Organ Failure Assessment Cardiovascular	0.019999	Sequential Organ Failure Assessment Cardiovascular	0.020905
2	Lactic Acid	0.013834	Discharge Location	0.012981	Hemoglobin	0.016193	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	0.014453	Sodium (135-148)	0.015098	Arterial pH	0.015704	Acute Physiology Score III	0.017487
3	**Arterial BP Mean**	0.013706	Acute Physiology Scores III Prob	0.012416	Hematocrit	0.014166	Elixhauser Ahrq Score Elixhauser Sid30	0.013737	SaO2	0.013504	Discharge Location	0.014116	Discharge Location	0.016266
4	**Respiration Rate (Total)**	0.013666	Simplified Acute Physiology Score	0.012075	Potassium (3.5-5.3)	0.014092	Lactic Acid	0.013733	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	0.013459	Age	0.013955	Potassium (3.5-5.3)	0.015801
5	Hemoglobin	0.013493	Lactic Acid	0.01169	Discharge Location	0.013913	Compliance (40-60ml)	0.012708	**NBP [Systolic]**	0.012678	Carbon Dioxide	0.0138	Arterial pH	0.015368
6	**Arterial BP [Systolic]**	0.013334	Acute Physiology Score III UO Score	0.011583	Magnesium (1.6-2.6)	0.013342	Respiration Rate (Total)	0.011896	Arterial PaCO2	0.012572	**Respiratory Rate**	0.013629	Age	0.014679
7	Sequential Organ Failure Assessment Cardiovascular	0.012407	Sequential Organ Failure Assessment	0.011426	**Arterial BP Mean**	0.01334	Elixhauser Ahrq Score Elixhauser Vanwalraven	0.011875	Compliance (40-60ml)	0.012195	BUN (6-20)	0.013144	Sequential Organ Failure Assessment	0.013576
8	RBC	0.012289	Age	0.011174	Age	0.013095	Sequential Organ Failure Assessment Cardiovascular	0.011855	Simplified Acute Physiology Score	0.011821	Organ Dysfunction (Y/N)	0.0124	Lactic Acid	0.012756
9	Arterial CO2(Calc)	0.011911	Tidal Volume (Obser)	0.010736	Compliance (40-60ml)	0.012857	Discharge Location	0.011746	WBC	0.011646	Acute Physiology Score III	0.012369	Creatinine (0-1.3)	0.01204
10	Elixhauser Ahrq Score Elixhauser Sid29	0.011749	Request MRSA Status Periodically (Y/N)	0.010714	Chloride (100-112)	0.012847	Acute Physiology Scores IIIProb	0.011549	Arterial PaO2	0.011629	Acute Physiology Scores IIIProb	0.012058	**Heart Rate**	0.011889

2.4.3 Case Study 2: Pediatric ICU Database

Data Source – CHOA

Our second case study uses de-identified data from Children's Healthcare of Atlanta containing 5000 patient records spanning an 11 month period. Each ICU stay record consists of the patient's demographic information (e.g., gender and age of admission), diagnosis (e.g., ICD-9 codes), birth related events (e.g., birth weight, head circumference, gestation weeks), microbiology events (e.g., microbes in blood or serum),. chart events (e.g., heart rate), medication intake events, microbiology events (e.g., microbes), and clinical records (e.g., heart rate, oxygenation) collected from bedside monitors, averaged over each min. A more comprehensive measures and number of records available from is shown in Table 2.8.

Table 2.8 Data Types in CHOA database

Data Type	Examples of Measures
Demographics	DOB, Gender, Age, Height, Weight, Ethnicity, Religion, Date of Death, Co morbidity with other diseases
Microbiology	Types of microbes, Amount of microbes, dilution
Lab Data	No of test performance, abnormalities in tests such as Urea, Albumin, Bilirubin, Creatinine, Sodium, Potassium, Calcium
Clinical Data	HR, Heart rhythm, BP, NBP, CVP, SaO_2, Arterial PH, Arterial $PaCO_2$, Arterial PaO_2, Arterial CO_2, SpO_2, Respiratory rate, Tidal volume, Respiratory effort, Hematocrit, WBC, RBC,
Medication Data	Medication & IV administered, Dosage, Duration time, Concentrations & Rate of Administration, composition of IV imposed

In this case study, we have data consisting of demographics, microbiology, diagnosis codes and medication data from which we extracted 9080 features. The missing data in this type is 0.08%, hence we used this only for decision making and not for missing data analysis. In addition to this we used lab data with a median sampling interval of 2.05

hrs. from which we extracted 2500 features and 1 min averaged vital signs data with a sampling interval of 1 min, from we extracted 44 features. The lab data consisted of information on the tests conducted (labeled as component name), the source of specimens (e.g. blood serum, urine and labeled as source), and the number of abnormalities in tests and procedures performed (labelled as Result status). The vital signs data consisted on 1 min average of the actual values. For the missing data analysis, we treat the lab data and vital signs data as 2 independent entities due to the varied temporal resolution.

Non-Temporal Data Analysis

For non-temporal analysis, we use the temporal data by converting it into mean values. Then features with greater than 80% missing data are removed. This gave us 1882 lab features and 44 vital signs features. Then outliers whose values which deviated by ±3 standard deviations from the mean value were removed.

Temporal Data Analysis

For non-temporal analysis, we chose the binning interval to be larger (4 hours for lab data and 5 min for wave data) than the median sampling interval. In addition, this dataset had an issue where the tests or values were not recorded for very long time intervals (~ days) in the middle. For this we treated the data as multiple time-series for each patient and did not interpolate the missing data for such large gaps. As a result, we did not encounter sampling related missing data in this dataset. In addition, some of the data for patients fell outside the range of the visit, in which circumstance we disregarded that data as it was most likely mislabeled.

Using a 2 hr. binned interval for lab data we the total missing data in the lab dataset was about 65% (mean) and a standard deviation of 2 and that for the vital signs data was 10% (mean) and a standard deviation of 18. Then we handled the missing data as stated above.

Test for "Neglectable" Assumption

We performed the t-tests and Little's test. The results of the t-test (Figure 2.17) and

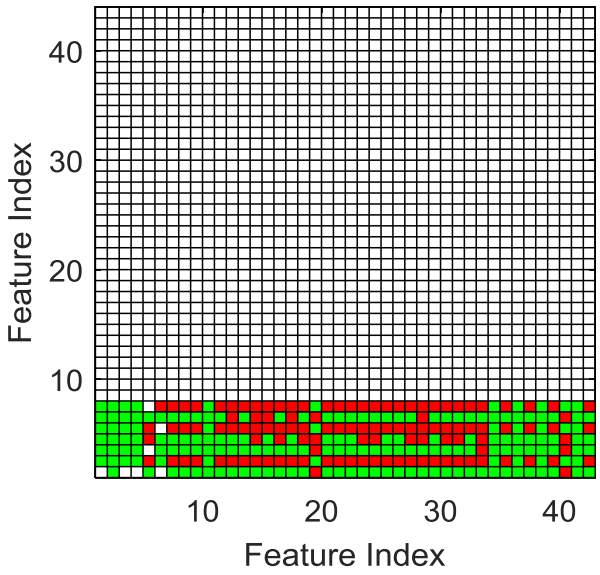


Figure 2.17: t-test results temporal vital signs: Green row means that particular feature is "Neglectable". No color means no missing data in feature or the missing data pattern is same as test feature. Since no row is green, data is not "Neglectable"

Little's (chi-square value 5261.17, degree of freedom of 1943 with a p-value 0.0) from non-temporal analysis of vital signs data demonstrates that the CHOA vital signs data is not "Neglectable." It is supported by the results of the temporal analysis which also proved that the vital signs data is not "Neglectable" (Figure 2.17). On Lab data also both non-temporal (Figure 2.18) and temporal analysis (Figure 2.19) suggest data is not "Neglectable".

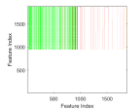


Figure 2.18: t-test results non-temporal lab: Green row means that particular feature is “Neglectable”. No color means no missing data in feature or the missing data pattern is same as test feature. Since no row is green, data is not “Neglectable”.

Identifying “NER” Data

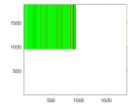


Figure 2.19: t-test results temporal lab: Green row means that particular feature is “Neglectable”. No color means no missing data in feature or the missing data pattern is same as test feature. Since no row is green, data is not “Neglectable”.

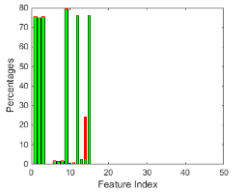


Figure 2.21: “NER” data identification vital signs non-temporal: The red bar gives percentage of all missing data the green bars give the “NER” data percentage. The features which have no missing data do not have any bars.

The classification analysis shows very high levels of the missing data to be “NER.” for both vital signs for non-temporal analysis (8% of the data and 95% of all missing data,

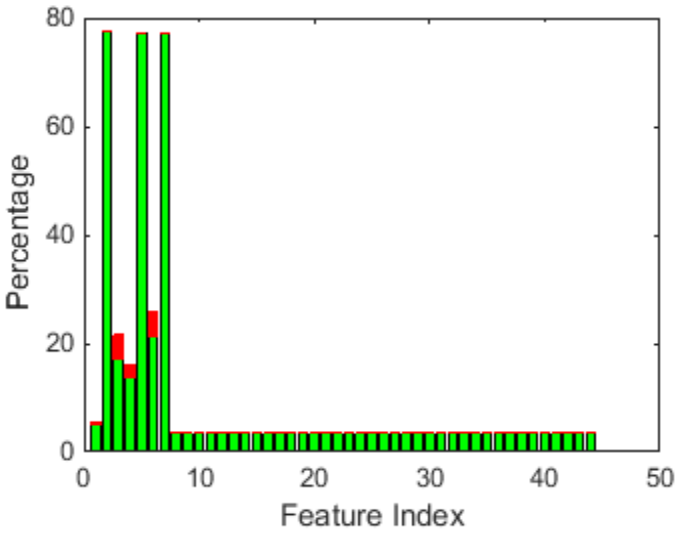


Figure 2.20: “NER” data identification vital signs temporal: The red bar gives percentage of all missing data the green bars give the “NER” data percentage. The features which have no missing data do not have any bars.

Figure 2.20) and for temporal analysis (9.5% of the data and 97% of all missing data, Figure 2.21). Similarly, lab data also showed high “NER” for both non-temporal (32% of the data

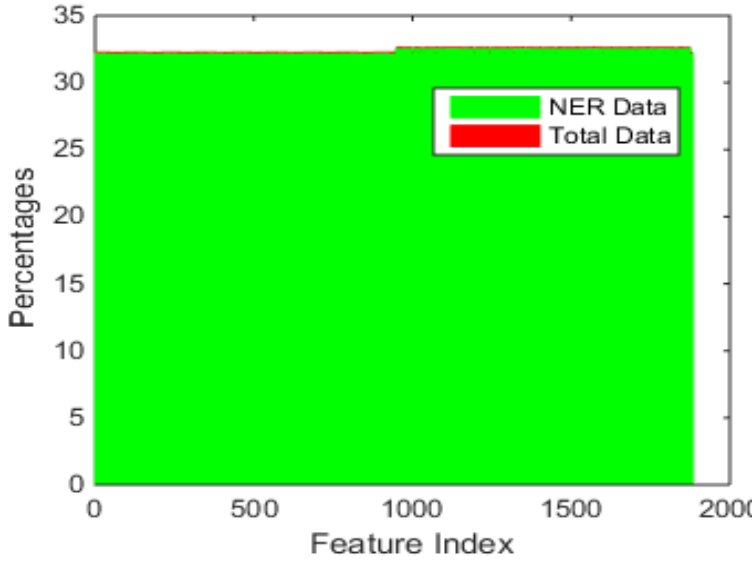


Figure 2.22: “NER” data identification lab non-temporal: The red bar gives percentage of all missing data the green bars give the “NER” data percentage. The features which have no missing data do not have any bars.

and 99% of all missing data, Figure 2.22) and temporal (65% of the data and 99% of all missing data, Figure 2.23)

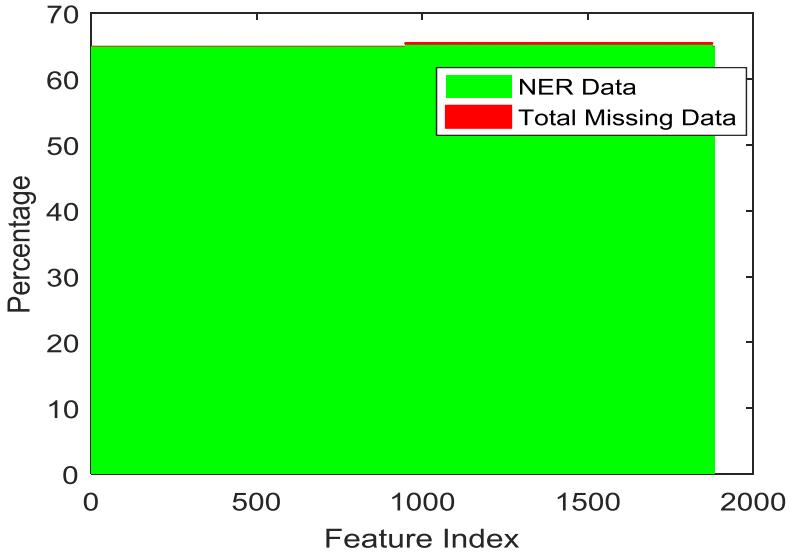
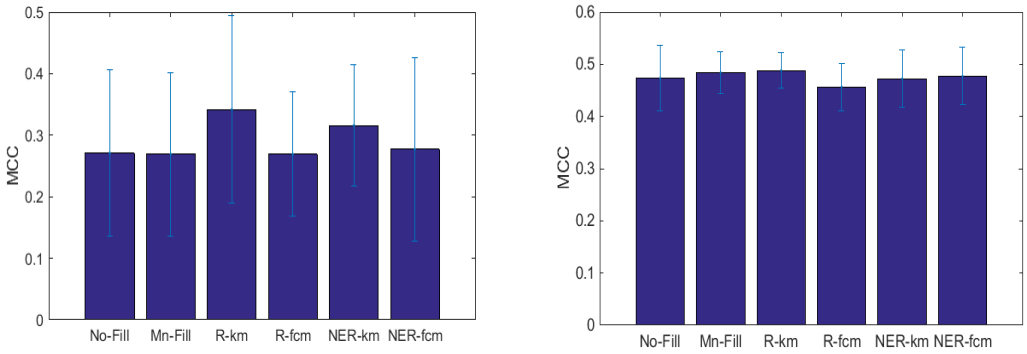


Figure 2.23: “NER” data identification lab temporal: The red bar gives percentage of all missing data the green bars give the “NER” data percentage. The features which have no missing data do not have any bars.

Evaluation using Random Forests

Imputation models were evaluated using Random Forests to predict ICU mortality for both vital signs and Lab data. The k-means based "Recoverable" models for Lab data outperformed both no filling and mean filling but the results were not significant at $p \leq 0.01$. The "NER" imputation (kmeans and fcm) were comparable to no-filling and mean filling (Figure 2.24a). For the vital signs data, the all the novel models outperformed the conventional models (Figure 2.24b). Both imputations with kmeans clustering (the "NER" and "Recoverable") were found to be statistically significant ($p \leq 0.01$) when the MCCs were compared using Steigler's Z-score [151].

The top repeated features (Tables 2.9, 2.10) for lab tests included the number of times of several procedures such as asoti, bilirubin test, biotin test, and for vital signs, it includes features such as respiratory rate, blood pressures, pulse, SpO_2, heart rates, sustained ventricular tachycardia, and abnormal ventricular rhythms.



a **b**

Figure 2.24: Mortality prediction results - MCC (fcm = fuzzy c means, km = kmeans, EM = expectation maximization, R = "Recoverable", NER = "Not-Easily-Recoverable" imputation). Figure a gives the results from using lab tests and the figure b gives the results from vital sign data.

Table 2.9: Top 10 CHOA- vital signs mortality: Features in each of the models with the importance scores and the number of features accounting for 90% of the total importance. All the features in the order of importance is given in the Appendix. The vital signs data was the actual values of the features such as respiratory rate, heart rate.

SN	No Fill	Imp 90% features = 19	Mn Fill	Imp 90% features = 24	EM	Imp 90% features = 20	kmeans MAR	Imp 90% features = 21	fcm MAR	Imp 90% features = 21	kmeans NER	Imp 90% features = 21	fcm NER	Imp 90% features =20
1	NBP diastolic	0.096272	NBP mean	0.079889	Respiratory Rate	0.064167	NBP diastolic	0.098396	Respiratory Rate	0.086222	Respiratory Rate	0.082667	NBP diastolic	0.074715
2	Respiratory Rate	0.092506	SpO2	0.071062	NBP mean	0.063523	Respiratory Rate	0.087994	Pulse	0.077041	NBP Systolic	0.071439	NBP mean	0.073837
3	SpO2	0.070491	NBP diastolic	0.063625	Arterial BP Diastolic	0.061245	Pulse	0.079265	NBP mean	0.076659	SpO2	0.060852	Pulse	0.071969
4	Run PVCs	0.060864	Heart Rate	0.061002	Heart Rate	0.059436	Arterial BP Diastolic	0.074551	Heart Rate	0.069704	Pacer Nt Pacing	0.060244	Respiratory Rate	0.064594
5	Heart Rate	0.060857	Respiratory Rate	0.057561	SD NN	0.058923	ABPs	0.055815	NBP Systolic	0.069037	Pulse	0.058792	Pacer Nt Pacing	0.064351
6	NBP mean	0.059527	NBP Systolic	0.057223	ABPs	0.058852	Heart Rate	0.055479	SpO2	0.066997	NBP diastolic	0.054042	NBP Systolic	0.063731
7	Pulse	0.047992	Pulse	0.052396	NBP Systolic	0.056301	NBP Systolic	0.052958	NBP diastolic	0.059302	Multiform PVCs	0.053332	SpO2	0.061171
8	NBP Systolic	0.046869	Run PVCs	0.045448	Arterial BP Mean	0.055101	Arterial BP Mean	0.050029	Pair PVCs	0.051877	Pair PVCs	0.048108	Pair PVCs	0.050922
9	Arterial BP Diastolic	0.040868	Frequent SVPBs	0.043225	Pulse	0.053171	SpO2	0.043269	SV Beats	0.047524	Frequent SVPBs	0.047894	Heart Rate	0.049752
10	Pause	0.039853	Multiform PVCs	0.035047	pNN50	0.048162	NBP mean	0.041497	Run PVCs	0.043633	pNN50	0.045374	Run PVCs	0.048648

Table 2.10: Top 10 CHOA- Lab Mortality: Features in each of the models with the importance scores and the number of features accounting for 90% of the total importance. The lab data consisted of information on the tests and procedures conducted (labeled as component name along with the procedure name or the just the test name), the source of specimens (e.g. blood serum, urine and labeled as source), and the number of abnormalities in tests and procedures performed (labeled as Result status)

SN	No Fill	Imp 90% features = 1815	Mn Fill	Imp 90% features = 1818	kmeans MAR	Imp 90% features = 1836	fcm MAR	Imp 90% features = 1832	kmeans NER	Imp 90% features = 1818	fcm MNAR	Imp 90% features =1810
1	COMPONENT NAME ALT (SGPT)	0.004773	COMPONENT STATUS Final	0.005298	SOURCE SYSTEM PROC CODE APG3	0.006442	COMPONENT NAME PATIENT FI02	0.006866	COMPONENT NAME Arterial POC PO2	0.005712	COMPONENT NAME Arterial POC PH	0.004681
2	PROC CATEGORY LAB - Blood GASES	0.004679	COMPONENT STATUS Corrected	0.005044	COMPONENT NAME FIBRINOGEN	0.006325	COMPONENT NAME ART BASE DEFICIT	0.006682	COMPONENT NAME BAND	0.005555	RESULT STATUS Edited	0.004596
3	COMPONENT NAME PROTIME	0.004465	COMPONENT NAME META	0.004925	COMPONENT NAME INT NORM RATIO	0.006091	COMPONENT NAME POC Potassium	0.006443	COMPONENT NAME WBC	0.005496	COMPONENT NAME META	0.004554
4	COMPONENT LINE 8	0.004638	COMPONENT NAME SEG	0.004925	COMPONENT NAME META	0.005914	SOURCE SYSTEM PROC CODE DICSCR	0.006185	COMPONENT NAME META	0.005342	COMPONENT LINE 8	0.004536
5	COMPONENT NAME Glucose	0.004509	COMPONENT NAME ART BASE DEFICIT	0.004906	COMPONENT NAME BAND	0.005762	COMPONENT NAME AUTOMATED ABS NEUT	0.006163	COMPONENT NAME ART BASE DEFICIT	0.004854	COMPONENT LINE 7	0.004471
6	RESULT HAS ABNORMAL CMPNT YN	0.0045	COMPONENT NAME LYMPH	0.004718	COMPONENT NAME % O2 SAT VENOUS	0.005725	COMPONENT LINE 4	0.006057	SOURCE SYSTEM PROC CODE APG3	0.004767	COMPONENT NAME VENOUS BASE EXCESS	0.004435
7	COMPONENT NAME C-REACTIVE PROTEIN	0.004481	COMPONENT NAME PHOSPHOROUS	0.004353	COMPONENT NAME ART BASE DEFICIT	0.005631	COMPONENT LINE 5	0.006021	SOURCE SYSTEM PROC CODE POCAI	0.004758	VALUE FLAG High Panic	0.004357
8	COMPONENT NAME NRBC	0.004458	COMPONENT NAME RBCS	0.004435	COMPONENT NAME Arterial POC PO2	0.00543	COMPONENT NAME MEAN PLT VOLUME	0.005979	VALUE IN RANGE YN	0.004722	COMPONENT NAME MYELO	0.004343
9	COMPONENT LINE 4	0.004457	COMPONENT NAME AST (SGOT)	0.004349	COMPONENT NAME POC ACT COAG TIME	0.00539	COMPONENT NAME Arterial POC PH	0.005965	COMPONENT NAME TOTAL PROTEIN	0.004666	COMPONENT NAME ALT (SGPT)	0.004288
10	COMPONENT NAME AST (SGOT)	0.004366	COMPONENT NAME BAND	0.004309	COMPONENT NAME DDIMER UNITS	0.005366	COMPONENT NAME CALCIUM	0.005961	RESULT STATUS Final result	0.00465	COMPONENT NAME Glucose	0.004288

2.4.4 Results Discussion

Table 2.11: Top 10 Features for Mortality for Best Performing Models. a. Features from MIMIC-II NER; b. Features from MIMIC-III NER; c. Features from CHOA vital signs Rec; d. Features from CHOA lab Rec;

Kmeans NER	fcm NER
Sequential Organ Failure Assessment Min	Hospital Length of Stay
Hospital Length of Stay	Simplified Acute Physiology Score I Max
SpO2	Sequential Organ Failure Assessment Max
Icustay Length of Stay	Sequential Organ Failure Assessment Min
Sequential Organ Failure Assessment Max	SpO2
Simplified Acute Physiology Score I Min	Simplified Acute Physiology Score I Min
Simplified Acute Physiology Score I Max	Cost Weight
Simplified Acute Physiology Score I First	Icustay Length of Stay
Cost Weight	Sequential Organ Failure Assessment First
Heart Rate	Heart Rate

a

kmeans NER	fcm NER
Discharge Location	Discharge Location
Arterial BP [Systolic]	Arterial BP [Systolic]
Arterial BP Mean	Arterial BP Mean
Sequential Organ Failure Assessment Liver	Sequential Organ Failure Assessment Liver
Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score	Acute Physiology Score III PaO2 Alveolar-Arterial Oxygen Gradient Score
Simplified Acute Physiology Score II	Simplified Acute Physiology Score II
Respiration Rate (Total)	Respiration Rate (Total)
Albumin (>3.2)	Albumin (>3.2)
Acute Physiology Score III Bilirubin Score	Acute Physiology Score III Bilirubin Score
Weight First	Weight First

b

kmeans Rec	fcm Rec
NBP diastolic	Respiratory Rate
Respiratory Rate	Pulse
Pulse	NBP mean
Arterial BP Diastolic	Heart Rate
ABPs	NBP Systolic
Heart Rate	SpO2
NBP Systolic	NBP diastolic
Arterial BP Mean	Pair PVCs
SpO2	SV Beats
NBP mean	Run PVCs

c

kmeans Rec	fcm Rec
SOURCE SYSTEM PROC CODE APG3	COMPONENT NAME PATIENT FI02
COMPONENT NAME FIBRINOGEN	COMPONENT NAME ART BASE DEFICIT
COMPONENT NAME INT NORM RATIO	COMPONENT NAME POC Potassium
COMPONENT NAME META	SOURCE SYSTEM PROC CODE DICSCR
COMPONENT NAME BAND	COMPONENT NAME AUTOMATED ABS NEUT
COMPONENT NAME % O2 SAT VENOUS	COMPONENT LINE 4
COMPONENT NAME ART BASE DEFICIT	COMPONENT LINE 5
COMPONENT NAME Arterial POC PO2	COMPONENT NAME MEAN PLT VOLUME
COMPONENT NAME POC ACT COAG TIME	COMPONENT NAME Arterial POC PH
COMPONENT NAME DDIMER UNITS	COMPONENT NAME CALCIUM

d

For mortality, we performed the analysis on MIMIC-II, MIMIC-III and CHOA datasets. In Table 2.11, we show the features from the best performing models in all the three datasets. When we compared the features, we found the common across all the

datasets were blood pressure, respiration rate, heart rate, pulse oximetry, albumin, creatinine, BUN,WBC and RBC abnormalities. All of these were non-specific ICU features. In addition, MIMIC-III and MIMIC-II showed cohort specific features such as SAPS and SOFA risk scores.

For mortality prediction, we also compared features from the MIMIC with that of the CHOA data. We found that features such as respiration rate, heart rate, pulse oximetry, arterial blood pressure and non-invasive blood pressure were top features in both datasets. Features such as height, weight, hematocrit tests, lactic acid, magnesium were on the bottom in both. Where features such as RBC, BUN, creatinine were in the middle on MIMIC-II features but remained in the top 10% on the CHOA. Certain comorbidities and tests for infection were at the bottom on both datasets.

For Sepsis, we formed the analysis only on MIMIC-II and MIMIC-III datasets, due to small number of sepsis patients in our CHOA dataset. In Table 2.12, we show the

Table 2.12: Top 10 Features for Sepsis for Best Performing Models. a. Features from MIMIC-II NER; b. Features from MIMIC-III NER;

kmeans NER	**fcm NER**
Sequential Organ Failure Assessment First	Sequential Organ Failure Assessment Max
Sequential Organ Failure Assessment Max	Hospital Length of Stay
Cost Weight	Cost Weight
Hospital Length of Stay	Sequential Organ Failure Assessment First
Heart Rate	Heart Rate
NBP	ICU Stay Admit Age
Respiratory Rate	Fluid Electrolyte Amount
Total Number of Hospital Stays	Simplified Acute Physiology Score I First
Fluid Electrolyte Amount	Lymphoma (Y/N)
Simplified Acute Physiology Score I Max	ICU Stay Length of Stay

a

kmeans NER	**fcm NER**
Sequential Organ Failure Assessment Cardiovascular	Sequential Organ Failure Assessment Cardiovascular
Arterial pH	Acute Physiology Score III
Discharge Location	Discharge Location
Age	Potassium (3.5-5.3)
Carbon Dioxide	Arterial pH
Respiratory Rate	Age
BUN (6-20)	Sequential Organ Failure Assessment
Organ Dysfunction (Y/N)	Lactic Acid
Acute Physiology Score III	Creatinine (0-1.3)
Acute Physiology Scores IIIProb	Heart Rate

b

features from the our novel methods in the two MIMIC datasets. When we compared the features, we found the common across all the datasets were non-specific ICU features such as respiration rate, heart rate, age, creatinine, and BUN abnormalities. In addition, they also showed cohort specific features such as SAPS and SOFA risk scores.

These results allow us to conclude that there are some cohort specific and general ICU features which allow the prediction of the endpoints such as mortality which can be generalized across datasets. In order to further test the generalizability of the different machine learning algorithms developed, we should use the models trained on the adult data and test on the pediatric data and vice-versa. In our current datasets, the number and type of data between the MIMIC and CHOA datasets is vastly different (mainly due to the issues of HIPPA and data availability). In the future, it will interesting to design a study where the same features are collected across different cohorts and the direct applicability/ generalizability of our algorithms is tested. This also high-lights a limitation of our current study, which requires retraining of the models and algorithms when the dataset is changed vastly.

2.4.5 Sensitivity Analysis

There are multiple parts of the missing data imputation where the design choice could have been made differently. The figure 2.25 shows the different designs we used to test our models.. We performed all our sensitivity analysis using MIMIC-II dataset.

Sensitivity of the Batch Size and Feature Order in Little's Test for "Neglectable"

In order to check whether batch sizes have an effect on the results, we performed Little's test using batch sizes of 3, 5, 7, and 10. Similarly, we repeated the test on three random ordering of the features, in order to check the effect of feature order on the final results.

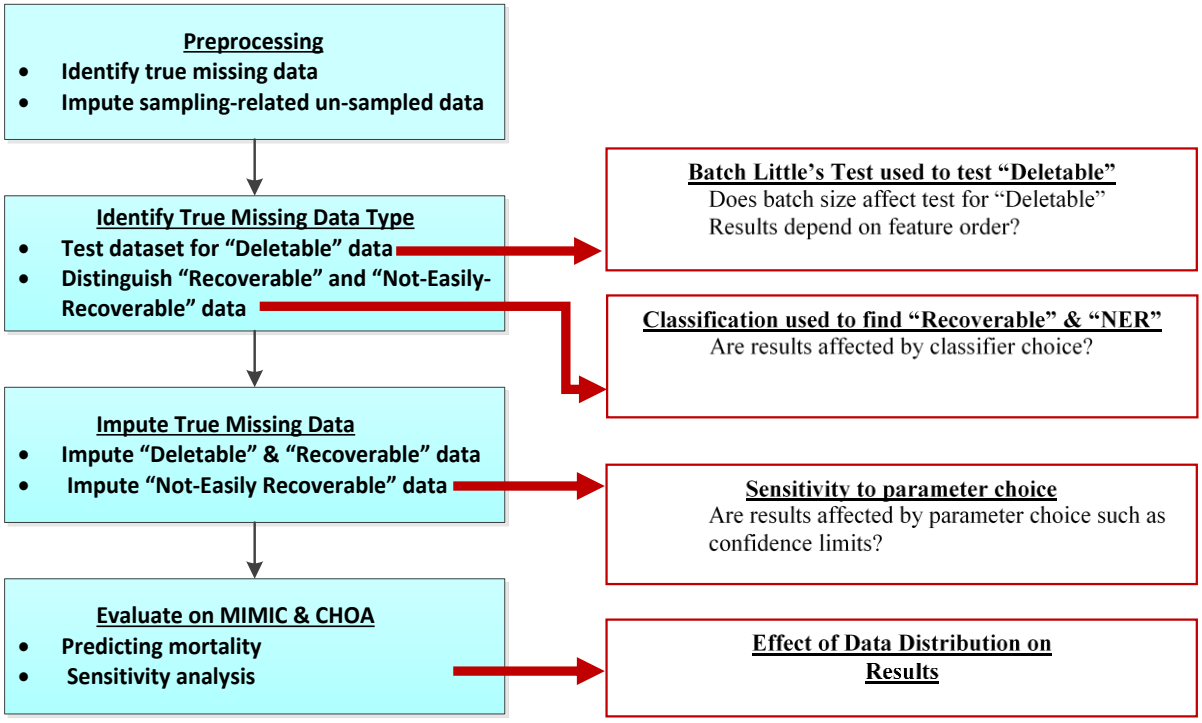


Figure 2.25: The various design choices used for which the sensitivity was tested.

Effect of Batch Size on the Result of "Neglectable" Test.

Table 2.13: Little's test results (Batch-Size = 3)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-3	3861.52	9	0
Feat 4-6	640.16	7	0
Feat 7-9	732.98	9	0
Feat 10-12	348.08	9	0
Feat 13-15	270.03	7	0
Feat 16-18	186.78	9	0
Feat 19-21	174.67	9	0
Feat 22-24	1466.96	9	0
Feat 25-27	366.07	9	0
Feat 28-30	55.6	7	0
Feat 31-33	485.05	9	0
Feat 34-39	3009.26	3	0
Feat 40-42	542.5	5	0
Feat 43-45	1902.69	3	0
Feat 46-48	392.96	3	0
Feat 47-87	1117.59	10	0

Table 2.14: Little's test results (Batch-Size = 5)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-5	6023.27	40	0
Feat 6-10	2062.79	75	0
Feat 11-15	1382.4	55	0
Feat 16-20	2275.72	32	0
Feat 21-25	1062.145	73	0
Feat 26-30	1767.2	53	0
Feat 31-35	409.22	46	0
Feat 36-40	2802.38	27	0
Feat 41-45	3020.52	4	0
Feat 46-50	3376.06	15	0
Feat 51-55	993.54	5	0
Feat 56-60	1617.11	11	0
Feat 61-65	7.134	1	0.007
Feat 66-87	12.549	3	0.005

Table 2.15: Little's test results (Batch-Size=7)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-7	7461.42	125	0
Feat 8-14	7228.7	212	0
Feat 15-21	2149.5	356	0
Feat 22-28	2340.69	143	0
Feat 29-35	903.67	62	0
Feat 36-42	3945.14	37	0
Feat 43-49	4358.51	21	0
Feat 50-87	43.32	6	0

Table 2.16: Little's test results (Batch-Size=10)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-10	12646.82	665	0
Feat 11-20	6928.14	725	0
Feat 21-30	3803.14	567	0
Feat 31-40	8278.82	77	0
Feat 41-50	6827.56	66	0

To test the effect of varying batch size on the iterative Little's test mentioned above, we tested the method using batch sizes 3, 5, 7 and 10. Results (Tables 2.13 - 2.16) indicate that the dataset is not "Neglectable" for all the different batch sizes and the results are independent of batch sizes. A possible reason for this is that if Little's test indicated

Table 2.17: Little's test (Random order 2)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-5	6023.27	40	0
Feat 6-10	2062.79	75	0
Feat 11-15	1382.4	55	0
Feat 16-20	2275.72	32	0
Feat 21-25	1062.145	73	0
Feat 26-30	1767.2	53	0
Feat 31-35	409.22	46	0
Feat 36-40	2802.38	27	0
Feat 41-45	3020.52	4	0
Feat 46-50	3376.06	15	0
Feat 51-55	993.54	5	0
Feat 56-60	1617.11	11	0
Feat 61-65	7.134	1	0.007
Feat 66-87	12.549	3	0.005

Table 2.18: Little's test (Random order 3)

Feature #	Chi Square	Degrees Freedom	P Value
Feat 1-35	101.19	3	0
Feat 36-40	1021.11	4	0
Feat 41-45	3257.49	8	0
Feat 46-50	532.79	75	0
Feat 51-55	528.60	75	0
Feat 56-60	2333.59	33	0
Feat 61-65	622.90	72	0
Feat 66-70	857.13	75	0
Feat 71-75	890.32	69	0
Feat 76-80	548.01	8	0
Feat 81-87	3924.14	50	0

"Neglectable" missing data, we combined multiple batches until all the data was used or some combination resulted is not "Neglectable".

Effect of Feature Order on the Result of "Neglectable" Test.

To test the effect of the feature ordering on the iterative Little's test mentioned above, we tested the method on 3 random ordering of the features in the dataset. Since we had concluded that the batch sizes did not affect the results based on the previous results we used a batch size of 5 for these tests. The results (Tables 2.17, 2.18) indicate that the dataset was not "Neglectable" in all of the orderings. The possible reason for this also is that the method is rerun until all data is used or some combination is not "Neglectable".

Sensitivity of the Classifier Choice for Distinguishing "NER" from "Recoverable"

The labels for training and classification was generated for each feature by assuming the value of 1 if data was missing and 0 otherwise. We tested neural networks, support vector machines (SVM), decision trees and LASSO L1 regularized logistic regression to distinguish "Recoverable" from "Not-Easily-Recoverable" since they all give a deterministic value each time. Any data that was missing and was labeled accurately was

considered to be missing ("Not-Easily-Recoverable"), and those which were mislabeled were considered to be imputable ("Recoverable"). This procedure was repeated for each of the different features. We report the correlation between the values of "Recoverable" and "NER" data evaluated using the different classification techniques.

The results from all the methods are very similar and show a correlation of greater than 0.9 (Table 2.19).

The results of classification analysis shows very high levels of the missing data to be "Not-Easily-Recoverable."(33.2% of the data and 99% of all missing data). These results indicate that most conventional approaches of imputing all the data using "Recoverable" assumptions or deleting may lead to bias.

Evaluation of the imputation models was performed using Random Forests to predict ICU mortality. The models where "Not-Easily-Recoverable" data was imputed using copulas outperformed those where "Not-Easily-Recoverable" data was not imputed. The fcm and k-means based "Recoverable" models outperformed traditional EM models (Table: 2.20 , 2.21), mean filling and no filling techniques. The MCC and accuracy of the novel models were similar irrespective of the classification technique used to distinguish the "Recoverable: from "NER" data. The statistical significance of these models was tested using Steigler's Z score [40] for correlated correlations from the MCC scores.

Table 2.19: Correlation between the "Recoverable" and "NER" data identified by the different classification techniques.

	LASSO	**Networks**	**Decision Trees**	**SVM**
LASSO	1.00	0.99	0.98	0.98
Networks	0.99	1.00	0.99	0.99
Decision Trees	0.98	0.99	1.00	0.98
SVM	0.98	0.99	0.98	1.00

Sensitivity of the Parameter Choice for "NER" Imputation

For the "NER" imputation we fit a multivariate copula under "NER" assumptions to sample from for estimating the "NER" data. In our study, we use a t-copula which is a function of the features and the "missingness" pattern R, (defined as 0 when a certain data is observed and 1 when otherwise). Each feature with missing data Y_i is then sampled from a distribution given by

$$Y_i \sim C(F_1(X_1), F_2(X_2), \dots, F_{\mathrm{N}}(X_N), F_{\mathrm{N+i}}(R_i)) \quad (2.3)$$

where $X = [X_1, X_2, \dots, X_N]$ is the data with N features and R_i is the missingness pattern for feature pattern for feature Y_i. The parameters for the copula are maximum likelihood

Table 2.20: MCC values of classification results of "Recoverable" and "NER" imputation with different methods to distinguish "Recoverable from "NER" data.

Conventional Methods			Recoverable Only		Recoverable + NER		
No Filling	Mean Filling	EM	kmeans	fcm	kmeans	fcm	
0.35 ± 0.03	0.39 ± 0.04	0.38 ± 0.05	0.4 ± 0.03	0.38 ± 0.03	0.54 ±	0.54 ± 0.01	LASSO
0.35 ± 0.03	0.39 ± 0.04	0.38 ± 0.05	0.38 ± 0.02	0.37 ± 0.04	0.54 ±	0.54 ± 0.02	Networks
0.35 ± 0.03	0.39 ± 0.04	0.38 ± 0.05	0.39 ± 0.03	0.36 ± 0.04	0.54 ±	0.54 ± 0.01	Decision
0.35 ± 0.03	0.39 ± 0.04	0.38 ± 0.05	0.38 ± 0.05	0.41 ± 0.04	0.54 ±	0.53 ± 0.02	SVM

Worst Performance Intermediate Performance Best Performance

Table 2.21: Accuracy values of classification results of "Recoverable" and "NER" imputation with different methods to distinguish "Recoverable from "NER" data.

Conventional Methods			Recoverable Only		Recoverable + NER		
No Filling	Mean Filling	EM	kmeans	fcm	kmeans	Fcm	
0.94 ± 0.003	0.939 ± 0.002	0.94 ± 0.005	0.94 ± 0.002	0.94 ± 0.002	0.95 ± 0.002	0.95 ± 0.001	LASSO
0.94 ± 0.003	0.939 ± 0.002	0.94 ± 0.005	0.94 ± 0.001	0.94 ± 0.002	0.95 ± 0.002	0.95 ± 0.002	Networks
0.94 ± 0.003	0.939 ± 0.002	0.94 ± 0.005	0.94 ± 0.002	0.94 ± 0.002	0.95 ± 0.001	0.95 ± 0.001	Decision Trees
0.94 ± 0.003	0.939 ± 0.002	0.94 ± 0.005	0.94 ± 0.002	0.94 ± 0.002	0.95 ± 0.001	0.95 ± 0.001	SVM

Worst Performance Intermediate Performance Best Performance

estimates fit using observed data and the "missingness" pattern at a p-value of 0.05. 0.10 and 0.15. The copulas with the different p-values all gave MCC values greater than 0.55 and accuracy greater than 0.95 (Table 2.22). All the MCCs were found to show an improvement over conventional techniques with a statistical significance of $p \leq .01$.

Effect of Population Imbalance on Missing Data Distribution

The dataset used in this study contains extremely unbalanced data, 2334 mortality records and 29,997 successful discharge records. The missing data was 30.6% ± 30.9% in patients with successful discharge from the ICU and 22.47% ± 30.44% in patients with ICU mortality. The correlation between the percentages of NER data in the two population was 0.98 and that of "Recoverable" data in the two populations was 0.87. These results indicate that the distribution of missing data in the two populations was similar.

Table 2.22: MCC and accuracy for NER imputation with different p-Value parameters

	MCC		Accuracy	
P-Value	**kmeans-NER**	**fcm-NER**	**kmeans-NER**	**fcm-NER**
0.05	0.54 ± 0.03	0.54 ± 0.01	0.95 ± 0.002	0.95 ± 0.001
0.1	0.56 ± 0.02	0.55 ± 0.02	0.95 ± 0.001	0.95 ± 0.002
0.15	0.55 ± 0.02	0.55 ± 0.01	0.95 ± 0.002	0.95 ± 0.001

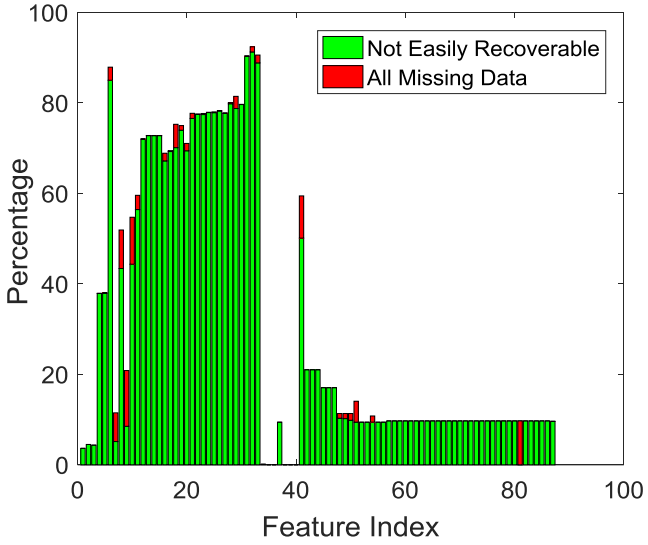


Figure 2.26: "Not-Easily-Recoverable" data identification in patients discharged: The red bar gives percentage of all missing data the green bars give the "Not-Easily-Recoverable" data percentage. The features which have no missing data do not have any bars in this figure.

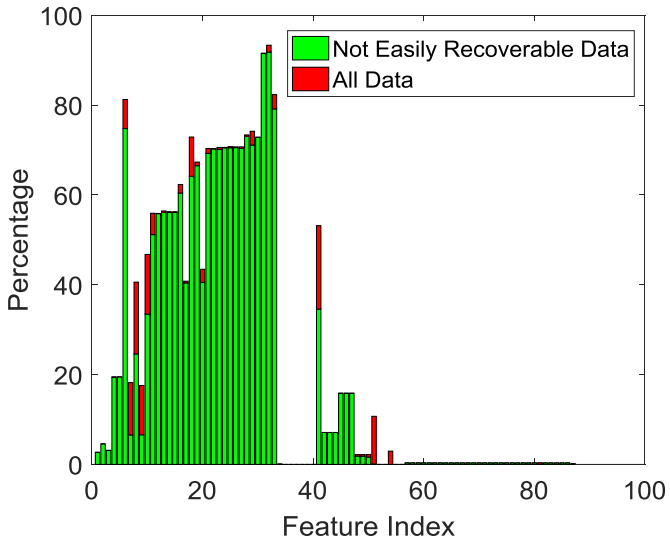


Figure 2.27: "Not-Easily-Recoverable" data identification in patients with ICU mortality: The red bar gives percentage of all missing data the green bars give the "Not-Easily-Recoverable" data percentage. The features which have no missing data do not have any bars in this figure.

The distribution of the different types of missing in the two populations are given in Figures 2.26, 2.27. To ensure that despite the imbalance in the populations is accounted when we compare the different imputation techniques, we use MCC as the evaluation technique. We chose MCC as the evaluation since its relatively insensitive to an imbalance in the population.

After performing the multiple sensitivity analyses, we conclude and our imputation technique is relatively robust to these design choices.

2.5. Summary and Key Innovations

The issues of data quality pose significant challenges to decision support systems. Conventional missing data interpolation and imputation schemes perform poorly because there are no models for modeling how the data is missing. In this study, we described the missing type into three categories namely "Neglectable", "Recoverable" and "NER". We demonstrated ICU data is not "Neglectable" and any deletion would result in bias. We then

proposed novel imputation for “Recoverable” and “NER” missing data types. We evaluated our results on two datasets consisting of adult ICU and pediatric ICU data, where our technique gave statistically significant ($p \leq .01$) improvement in performance as compared with EM, mean filling, and no missing data models. In the future the study evaluation can be expanded to use other endpoints such as ICU readmission and sepsis. Also for the “Recoverable” data imputation, we used kmeans and fcm for clustering. This can be expanded to include more robust of the clustering techniques such as genetic algorithms [140], the use of trimming procedures [169], hierarchical and density based clustering, in addition to particle swarm optimization [1] and missing interval size [2], which directly utilize patient data for clustering.

The key innovations of this chapter include:

- Categorization of and testing of missing data types in ICU
- Development of two novel methods for two kinds of missing data in EHR
- Evaluation on Adult and Pediatric datasets

CHAPTER III

TIME-SERIES DATA ANALYSIS TO PREDICT ADVERSE OUTCOMES IN THE INTENSIVE CARE UNIT

3.1. Introduction

After addressing the challenge of missing data in chapter 1, I used the two best performing models ("NER" imputation with kmeans and fcm) for predicting adverse ICU outcomes using data temporal analysis. Research on the analysis of patient data to predict adverse events such as ICU mortality, ICU readmission and sepsis have mainly used probabilistic models. Logistic regression [28, 33, 78], Cox regression [78] and artificial neural networks [28, 80, 170] are the most common models used in the analysis of healthcare data. However, these models suffer from inherent issues, particularly their basis on using a snapshot of the data available to make longitudinal predictions. These models often make use of a single time point to make predictions about subsequent adverse events occurring in the ICU including mortality and ICU readmission.

In this study, we address these issues by proposing a retrospective study of adult and pediatric ICU populations to discover factors indicative of adverse events such as ICU mortality and 30-day ICU readmissions using Conditional Random Field (CRF) models. CRFs are capable of making predictions of time series data by utilizing the parameters learned from a large patient population. They are well studied for sequential data such as natural language processing [171], structured prediction in computer vision, imaging [172-174], sleep studies [175, 176] and activity modeling [177-179]. CRFs have also been used for waveform analysis in health data [99-103] and are particularly advantageous in the ICU over the current models. In this study, we compare CRF with traditional techniques of

logistic regression and feed-forward artificial neural networks (NN) for ICU patient data. We also extend CRF by combining it with survival analysis to give the physicians an idea of the risk to a patient over time.

We structure the remainder of this chapter as follows. First, a short description of our data source is followed by a detailed description of the preprocessing and data mining approaches in section 2. Evaluation, results, and discussion are presented in section 3. Finally, the conclusion and key innovations are summarized in section 4.

3.2. Methods

In this study, we perform the classification of ICU patients into high risk and low risk for adverse events using a temporal mining technique called CRF. We demonstrate our results using a retrospective data analysis of adult and neonatal ICU data from Multi-parameter Intelligent Monitoring in Intensive Care, (MIMIC-II) database. We use CRF to determine patient's factors, which contribute to adverse consequences such as ICU mortality, and 30-day ICU readmission. These end-points are particularly interesting since they provide the basis for the long-term prediction of adverse events.

3.2.1 Data Pre-Processing

Each ICU stay record consists of the patient's demographic information, diagnosis, chart events, medication intake events, microbiology events etc. Each patient record consists of features which are either static (does not change over the entire duration of the patient's ICU stay) or temporal (changing in time). For this analysis, the data from each feature was binned using a fixed binning interval. The missing data was divided into the three types mentioned in Chapter 1 ("Neglectable", Recoverable" and "NER"). Then each type was imputed differently using the techniques described in Chapter 1 [180]. "NER"

data was imputed using student's t-copulas and "Recoverable" data was imputed using expectation maximization (EM) after clustering [180]. Both kmeans ("NER" kmeans) and fuzzy C means ("NER" fcm) were used for clustering the data prior to imputation here. We will refer to these two imputation techniques as 'Imp-1' and 'Imp-2'.

We then proceed to perform feature selection and classification on the patient data sequences for identifying patients at risk for adverse events such as mortality in the ICU and 30-day readmission

3.2.2 Feature Selection

The data often contains features, which may have low correlation with the outcome and sets of features may contain redundant information. In order to remove such features, we used L1 feature selection to keep the features space sparse. L1 feature selection typically penalizes the absolute values of the weights using a L1 regularization parameter. This parameter is set using 3×3 nested cross-validation technique [147].

3.2.3 Classification using Conditional Random Fields (CRF)

Following feature selection, we perform classification of ICU temporal data using CRF. CRFs are graphical models that encode the probability distribution ($p(\mathrm{y}|\mathrm{x})$) of a set of outcomes (y) given the features (x) (e.g. probability distribution of the likelihood of patients most likely to suffer adverse consequences such as ICU mortality and ICU readmission). It was first introduced by Lafferty et. al. [171] as an improvement over existing sequence classification methods such as Markov models because CRFs do not require assumptions of independence, and do not place any constraints on the distributions of the features and outcome variables. In addition, CRF overcomes the labeling bias (i.e.

for Markov models weight distribution is biased towards states with fewer successor states due to local weight space) by using a global weight space [171, 181, 182].

CRF models (Figure 3.1) directly represent the conditional probability of a particular label sequence,y $\in$ Y given a sequence of observations $x = \{x_1, x_2 \ldots x_T\}$ i.e. $p(y|x, \theta)$, where θ is the set of parameters, y is the outcome and x is the feature vector. Each of the observations is represented by a feature vector of dimensionality d ($x_i \in R^d$). Each hidden variables $h = \{h_1, h_2 \ldots h_T\}$ represent a higher order feature derived from the combination of features x. At any given time instance (t) the observations x_t connect to the nodes in the hidden states h_t that take a value from a finite set H. The probability for patient k $P(y^k, h|x^k, \theta)$ is given by equation. 3.1.

$$P(y^k, h|x^k, \theta) = \frac{1}{Z} e^{\left(\theta\varphi(y^k, h, x^k; \theta)\right)} \tag{3.1}$$

where θ is the set of parameters estimated during training, $\varphi(y^k, h, x^k; \theta)$ is the clique potential function parameterized by θ. A clique is a fully connected sub-graph and clique potentials are exponential functions of y^k and h in the clique [183]. Cliques in a chain CRF (used here) consists of an edge between adjacent hidden variables (h_{t-1} and h_t) and the edges from those two outcomes to the set of observations x^k and outcomes y^k.

As a result, CRFs represent the conditional probability as (equations 3.2-3.4):

$$P(y^k|x^k, \theta) = \sum_h \frac{1}{Z} e^{(\theta.\varphi(y^k, h, x^k; \theta))} \tag{3.2}$$

where,

$$Z = \sum_{y,h} e^{(\theta.\varphi(y^k, h, x^k; \theta))} \tag{3.3}$$

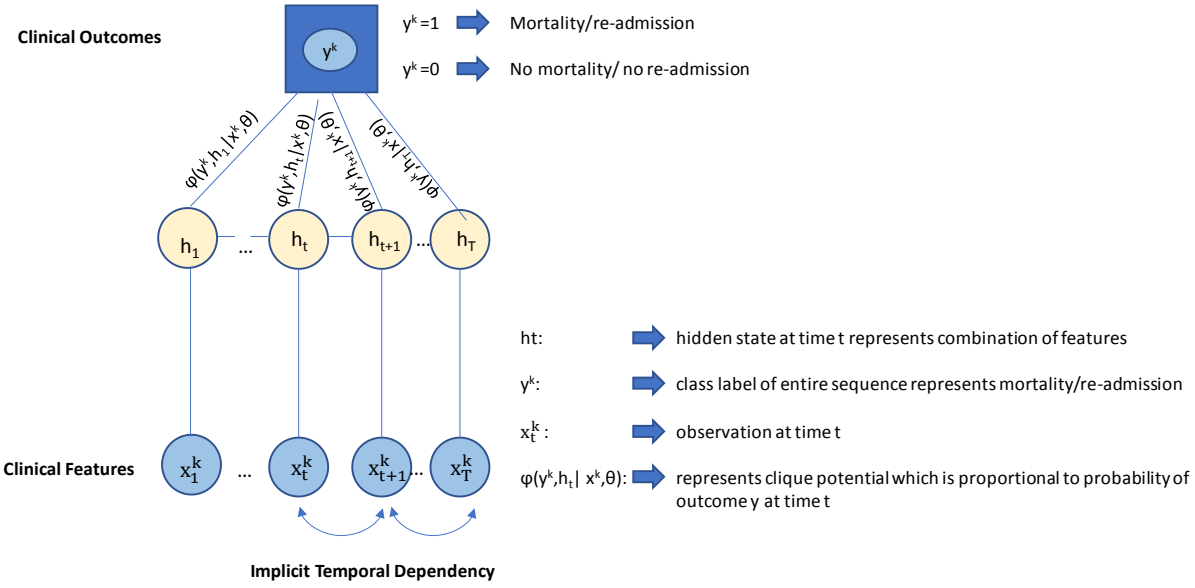


Figure 3.1: Linear-Chain Hidden Conditional Random Field Structure used to Predict Adverse Events in the ICU (30 day ICU Readmission and ICU Mortality).

$$\varphi(\mathrm{y}^k,\theta,h,\mathrm{x}^k;\theta)=\sum_{j=1}^{T}\sum_{l\in F} f_l^1(j,\mathrm{y}^k,h_j,h_{j-1},\mathrm{x}^k)\theta_l^1 \;+\; \sum_{j,k\in E}\sum_{l\in F} f_l^2(j,k,\mathrm{y}^k,h_j,h_k,\mathrm{x}^k)\theta_l^2 \tag{3.4}$$

where, E, F are the number of edges and features respectively. And f_l^1, f_l^2 are feature transformation functions (analogous to regression here)

Hence, the log-likelihood function is given by equation 3.5

$$P(\mathrm{y}^k|\mathrm{x}^k,\theta)=\log\left(\frac{1}{Z(X)}\times\sum_{\mathrm{h}}\frac{1}{\mathrm{Z}}\mathrm{e}^{\left(\theta.\varphi(\mathrm{y}^k,\mathrm{h},\mathrm{x}^k;\theta)\right)}\right) \tag{3.5}$$

The value is maximized to learn the parameters θ. The inference is done by forward-backward inference to obtain the outcome probability from the graph. Over-fitting of the CRF model is prevented using L1 regularization of weights (the absolute values of weights are penalized).

3.2.4 Extension of CRF with Survival Analysis

The temporal profile of the hazard faced by patients is essential for physician to adjust treatment. In this study (Figure 3.2), we extend the work by Lin *et. al.* [98] to combine CRF with survival curves to show the temporal risk profiles per patient.

As we know, the probability per patient is given by equation 3.6

$$P(\mathrm{y}^k|\mathrm{x}^k) = \frac{1}{Z(\mathrm{x}^k)} \times \sum_h exp(\varphi(\mathrm{x}^k, h, \mathrm{y}^k; \theta)) \tag{3.6}$$

which can be rewritten as equation 3.7

$$P(\mathrm{y}^k, \mathrm{h}|\mathrm{x}^k) = \frac{1}{\mathrm{Z}(\mathrm{x}^k)} \times \{\exp(\sum_{\mathrm{j}=1}^{\mathrm{T}} \sum_{\mathrm{l}\in\mathrm{F}} \mathrm{f}_\mathrm{l}^1(\mathrm{j}, \mathrm{y}^k, \mathrm{h}_\mathrm{j}, \mathrm{x}^k)\theta_\mathrm{l}^1 + \sum_{j,k\in E} \sum_{l\in F} f_l^2(j, k, \mathrm{y}^k, h_j, h_k, \mathrm{x}^k)\theta_l^2\} \tag{3.7}$$

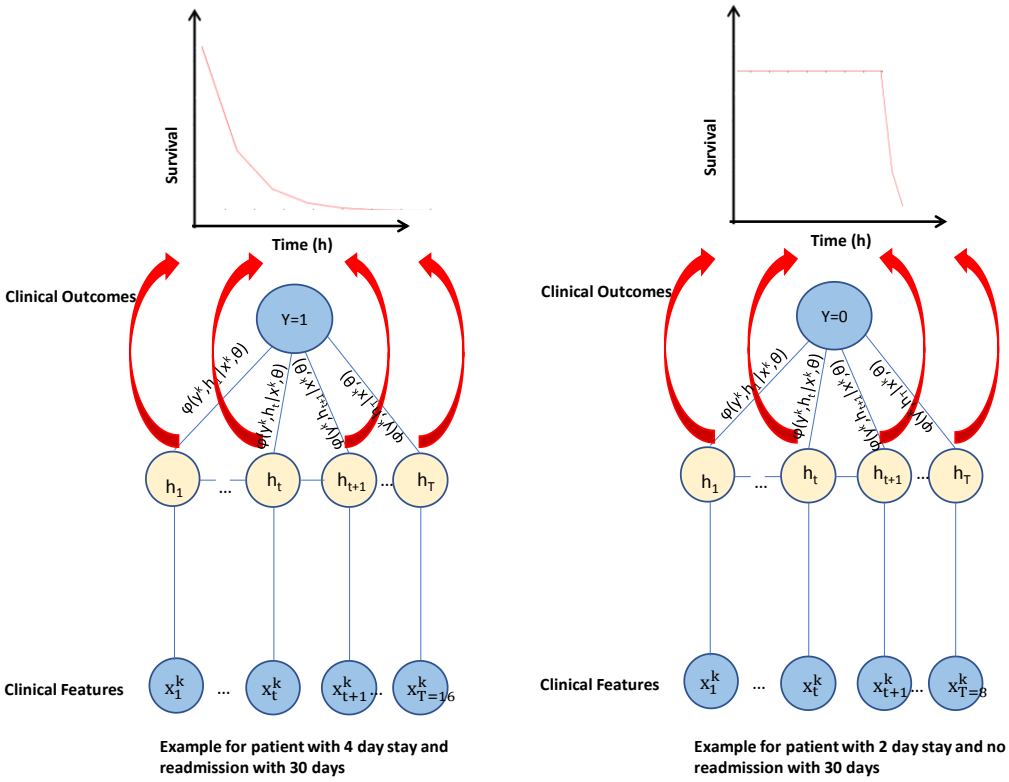


Figure 3.2: Incorporating Survival Analysis into CRF to Project the Temporal Patient Risk Profile by using P(yk,ht|xk,θ) as the Hazard function for Survival Analysis.

The time varying beliefs gives the temporal contribution to the likelihood of the final outcome. The time varying beliefs $P(y^k, h_t|\mathrm{x}^k, \theta)$ give the temporal contribution to the likelihood of the final outcome at time t. As shown in Fig. 1b, they are equivalent to the hazard function in survival analysis to denote the probability of occurrence (or risk) of the adverse event at the specific time instant t. Thus, the survival curve is calculated using the equation 3.8 [98]

$$S(t) = exp(-\int_0^t P(\mathrm{y}^k, h_u|\mathrm{x}^k, \theta)du) \tag{3.8}$$

In our study we develop a Matlab graphical user interface (GUI) to display these survival curves.

3.2.5 Hyper-Parameter optimization & Evaluation

In order to develop an efficient model, the hyper-parameters of the models must be estimated to allow for robust performance in a generalizable context. Hence the parameters to be optimized in training include the number of hidden variables and L1 regularization constant. The estimation of these hyper-parameters and evaluation of the models obtained is performed using 3×3 nested cross-validation technique [147].

The data is first split into two (training and validation dataset). The training set is further divided into testing and training set in the inner loop. The training-set in the inner loop is used to train the model, and the test set is used for parameter optimization. The inner loop is used to optimize the two parameters. Three-fold cross validation is performed with three repetitions and the average Matthew's correlation coefficient (MCC) and accuracies are computed on the test data. The number of hidden states ranges from 2 to 8, and the regularization parameter is varied exponentially from .01 to 100. The best performing model is chosen on the basis of best-averaged MCCs. The best performing

Table 3.1: Values used for Hyper-parameter Optimization for CRF, LR and NN

Model	Hyperparameter	Values
CRF	Number of hidden states	2,4,6,8
	Regularization Constant	1,3,10,30,100
NN	Number of hidden layers	2,4,6,8
	Number of Features	3,6,9,12,15,18,21,24,27,30
LR	Regularization Constant	0.01, 0.03, .1, 0.3, 1,3,10,30
	Number of Features	3,6,9,12,15,18,21,24,27,30

model is the validated against the validation set (Table 3.1). This whole process is repeated 3 times to get the average performance of the model.

3.2.6 Comparison of Existing Methods

CRF performance was compared with traditional models of logistic regression (LR) and feed forward neural networks (NN). LR uses L_1 regularization and Minimum Redundancy Maximum Relevance (mRMR) technique [184] to extract features. For L1 regression, the features with highest coefficients were considered as the selected features. NN uses mRMR only for feature ranking because there is no one-to-one correspondence of NN weights with the features. In CRF we used L1 regularization only because mRMR cannot be done without temporal labels. The number of features obtained from mRMR is optimized so that the classification model yields the best average performance on the validation set. For traditional methods like LR and NN, they do not take temporal patterns automatically. To perform a fair comparison of feature selection between CRF and LR/NN, we input an average value of the temporal pattern to LR and NN.

The evaluation metrics reported were Matthew's correlation coefficient (MCC), area under the curve (AUC) and accuracies computed on the test data. We also tested the CRF models for statistical significance using Steiger's Z score for correlated correlations [151].

3.3. Results & Discussion

We demonstrate our methods of patient classification using temporal data (CRF) and the use of individual risk profiles using survival curves on adult ICU populations from MIMIC-II data. We test our methods on the adult for the end-points ICU mortality and 30 day ICU readmissions. These end-points represent most common end-points for the respective populations in literature and pose great risks to the patients.

3.3.1 Adult ICU Database – MIMIC-II Database

This study is a retrospective data analysis using data from Multi-parameter Intelligent Monitoring in Intensive Care, second version, (MIMIC-II) database. MIMIC-II is a public ICU data repository containing over 40,000 ICU stay records (32,331 adult and 8080 neonatal records) [149]. The MIMIC II data for each patient is either static (does not change over the entire duration of the patient ICU stay, e.g., patient demographics) or temporal (changing in time, e.g., heart rate, blood pressure. A total of 87 features (33 temporal & 54 static features) which covered clinical measurement, lab results administrative data, comorbidities and other diagnostic procedures, were used for classification. The static features defined here are features which do not change during the duration of the stay (e.g. presence or absence of cancer, diabetes, age in years, etc.). Hence, the static features were repeated at each time instant for the CRF model.

After binning into intervals of 6 hours, data had 87± 21% missing data. The missing data was imputed as stated above.

3.3.2 Results to Predict ICU-Readmission

A total of 84 features (33 temporal & 51 static features) were used to classify patients at risk of 30 day ICU readmission. There are 7,787 patients having an ICU readmission within 30 days versus 24,544 without ICU readmission within 30 days.

The averaged values of MCC, AUC and accuracies is shown in Table 3.2. For each of the aforementioned models (CRF, LR, NN), we used two data pre-processing methods MNAR with kmeans (Imp-1) and MNAR with fuzzy-c-means (Imp-2).

Results (Table 3.2) indicate that the CRF Imp-2 (MCC = 0.73±0.03) models outperformed all LR and NN models with a statistical significance $p \leq 0.01$ when compared using Steiger's Z score for correlated correlations [151]. CRF Imp-1 model outperformed all LR models with a statistical significance $p \leq 0.01$. Comparing CRF Imp-1 with NN, the results were not statistically significant at $p \leq 0.01$. This shows that CRF results showed a statistically significant improvement at $p \leq 0.01$ over LR when using different data imputation techniques.

Table 3.2: Classification Results from ICU Readmission
(LR = Logistic regression, NN = Neural networks, CRF = Conditional random fields, Imp-1= MNAR kmeans, Imp-2 = MNAR fcm [1].)

	LR (L_1) Imp1	**LR (L_1) Imp2**	**LR (mRMR) Imp1**	**LR (mRMR) Imp2**	**NN (mRMR) Imp1**	**NN (mRMR) Imp2**	**CRF (L_1) Imp1**	**CRF (L_1) Imp2**
Accuracy	0.79± 0.003	0.79 ± 0.002	0.79 ± 0.001	0.79 ± 0.001	0.80 ± 0.0	0.80 ± 0.001	0.80 ± 0.006	0.90 ± 0.013
MCC	0.33 ± 0.014	0.33 ± 0.010	0.32 ± 0.005	0.32 ± 0.007	0.39 ± 0.004	0.38 ± 0.003	0.39 ± 0.021	0.73 ± 0.032
AUC	0.77 ± 0.002	0.77 ± 0.006	0.76 ± 0.002	0.76 ± 0.003	0.80 ± 0.007	0.80 ± 0.009	0.84 ±0.027	0.95 ± 0.015

(a) **(b)**

Figure 3.3: 30 day ICU readmission sensitivity plot giving the MCC values when two of most influential weights were perturbed using 50 values in the interval ± 10%.
(a) gives the sensitivity analysis for Imp-1 with MCC
(b) gives the sensitivity analysis for Imp-1 with MCC

We also ran a sensitivity analysis on our models by perturbing the model parameters by 10%, one at a time and found that there was no significant changes in the performance. Figure 3.3 shows the MCC values when two of most influential weights were perturbed using 50 values in the interval ± 10%. This further shows the robustness of the CRF models despite noisy data.

3.3.3 Results to Predict ICU-Mortality

For the end-point of ICU mortality, all the 87 features (33 temporal & 54 static features) were used. There are 2,334 patients who passed away during the ICU stay and 29,997 patients without mortality in the ICU.

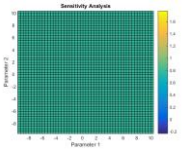
The averaged values of MCC, AUC and accuracies is shown in Table 3.3. Results indicate that the CRF Imp-1 (MCC =0.50 ± 0.033) models outperformed all LR models with a statistical significance $p \leq 0.01$. CRF Imp-2 model outperformed L_1 regularized LR models with a statistical significance $p \leq 0.01$. The results of CRF-Imp1 was comparable with the NN models

Table 3.3: Classification Results from ICU Mortality
(LR = Logistic regression, NN = Neural networks, CRF = Conditional random fields, Imp-1= MNAR kmeans, Imp-2 = MNAR fcm [1].)

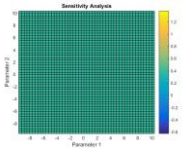
	LR (L_1) Imp1	**LR (L_1) Imp2**	**LR (mRMR) Imp1**	**LR (mRMR) Imp2**	**NN (mRMR) Imp1**	**NN (mRMR) Imp2**	**CRF (L_1) Imp1**	**CRF (L_1) Imp2**
Accuracy	0.94 ± 0.000	0.94 ± 0.001	0.94 ± 0.001	0.94 ± 0.001	0.94 ± 0.000	0.94 ± 0.001	0.94 ± 0.003	0.94 ± 0.004
MCC	0.42 ± 0.007	0.42 ± 0.023	0.47 ± 0.006	0.48 ± 0.009	0.51 ± 0.006	0.51 ± 0.007	0.50 ± 0.033	0.46 ± 0.098
AUC	0.90 ± 0.002	0.90 ± 0.006	0.90 ± 0.004	0.90 ± 0.005	0.90 ± 0.008	0.90 ± 0.001	0.88 ± 0.003	0.89 ± 0.011

We ran a sensitivity analysis on our mortality models by perturbing the model parameters by 10%, one at a time and found that there was no significant changes in the performance. Figure 3.5 shows the MCC values when two of most influential weights were perturbed using 50 values in the interval ± 10%. Similar to the results on ICU- readmission, ICU mortality showed a robustness and invariance to perturbations in model parameters.

So far our results indicate that CRF had better prediction as compared to LR for both end-points. It outperformed NN for 30-day ICU readmission and was comparable NN



(b)



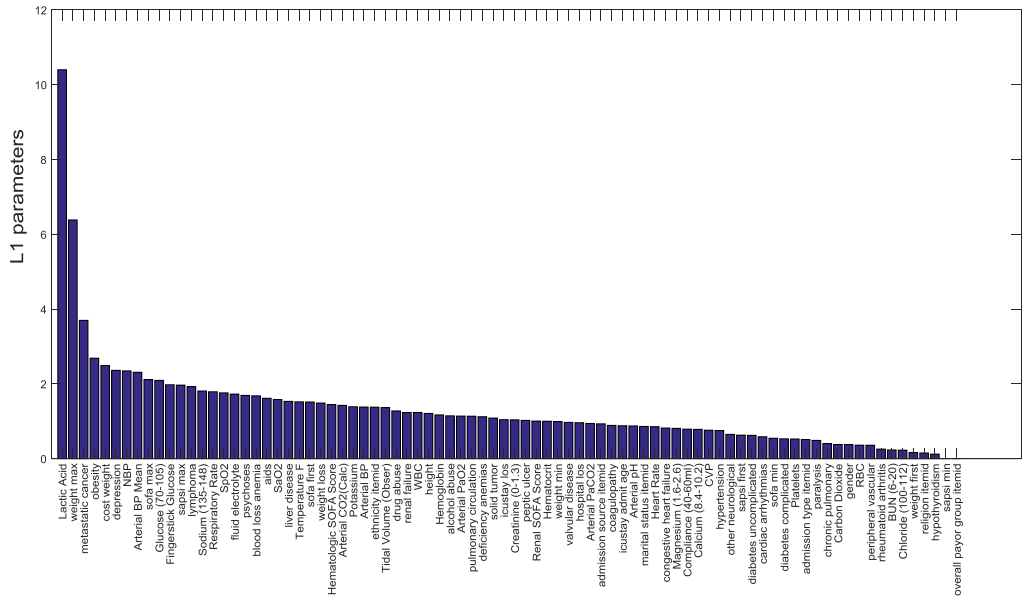
(b)

Figure 3.4: ICU mortality sensitivity plot giving the MCC values when two of most influential weights were perturbed using 50 values in the interval ± 10%.
(a) gives the sensitivity analysis for Imp-1 with MCC
(b) gives the sensitivity analysis for Imp-1 with MCC

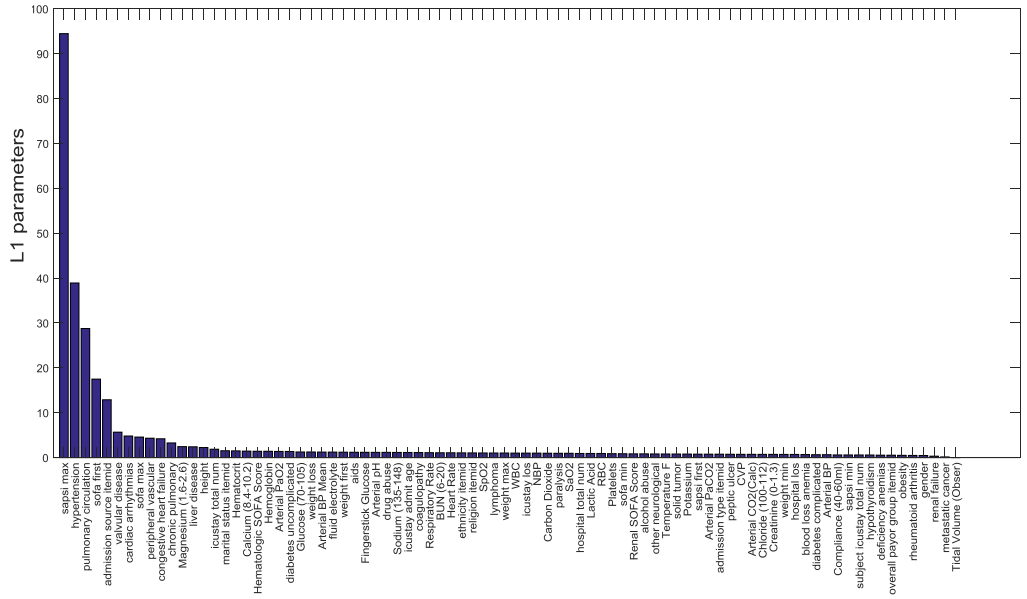
for mortality. In addition, the features picked using temporal models were different from those obtained from LR and NN. This leads us to conclude that the addition of temporal information gives different patterns in data which could provide more valuable insight into the disease processes. In the next aim, we investigate models which can combine the advantages and information content captured by both static models (LR, NN) and temporal models such as CRF. This type of foresight can help guide both the immediate management of a patient and the overall resource utilization.

3.3.4 Result Interpretation & Discussion

For each of the two endpoints ICU mortality and 30-day ICU readmission, we computed contributions of each of the features towards the final decision. For CRF models, the parameters of the models were indicative of the contribution of each feature towards the final decision. As mentioned above, we used L1 regularization for feature selection. Using the magnitude of parameters generated from the CRF models and L1 regularization, we found 90% of the features which contributed to the final decision making and the contribution of each of each of the features to the final decision. Similarly, for LR with L1 regularization, the parameters are the odds ratio and represent the contributions of each feature towards the decision (a positive value determines the contribution to the adverse event and a negative value determines the contributions to the other class). As with CRF used the magnitude of the parameters to determine which features contributed towards 90% of the final decision. For the LR and NN where we used mRMR the mutual information calculated was used to determine the relative contribution of each feature to the final decision making and we used that information to determine the features which contributed to 90% of the decision.



(a)



(b)

Figure 3.5: 30 day ICU readmission results showing the top features from CRF with L1 regularization-
(a) Gives a plot of mutual information with kmeans MNAR imputation i.e. Imp-1
(b) Gives a plot of mutual information with fcm MNAR imputation i.e. Imp-2

Results from 30 day ICU-readmission analysis (Figure 3.5, Table 3.4) indicate that for CRF with L1 regularization, for Imp-1, the physiological features and presence or

absence of disease contributed most to decision. A total of 59 features contributed to 90% of the decision. Similarly Imp-2 the Simplified Acute Physiology Score (I) (SAPS-I), physiological features and presence or absence of disease contributed most to decision with 43 features contributing to 90% of the decision. For LR with L1 regularization, for both Imp-1 and Imp-2, the top features (Figure 3.6, Table 3.4) which correlated with readmission were the hospital length of stay and physiological scores. The top feature which correlated with lack of readmission was the presence or absence of blood loss anemia. When the absolute value of the parameter values were used for computing the 90% contribution, the presence or absence of blood loss anemia contributed towards 90% of the distinguishing capacity between the two groups. When only the features which correlated with the presence of readmission was considered, a total of 14 for Imp-1 and 16 for Imp-2 features contributed to 90% of the classification. LR with mRMR (Figure 3.7, Table 3.4) the top features which contributed to classification were the factors such as the total number of ICU stays, followed by physiological parameters. For Imp-1, 20 features contributed to 90% of the classification and for Imp-2, 26 features contributed to 90% of the classification. For NN with mRMR (Figure 3.8, Table 3.4) the top features which contributed to classification were the number of ICU stays, followed by physiological parameters. For Imp-1, 20 features contributed to 90% of the classification and for Imp-2, 23 features contributed to 90% of the classification.

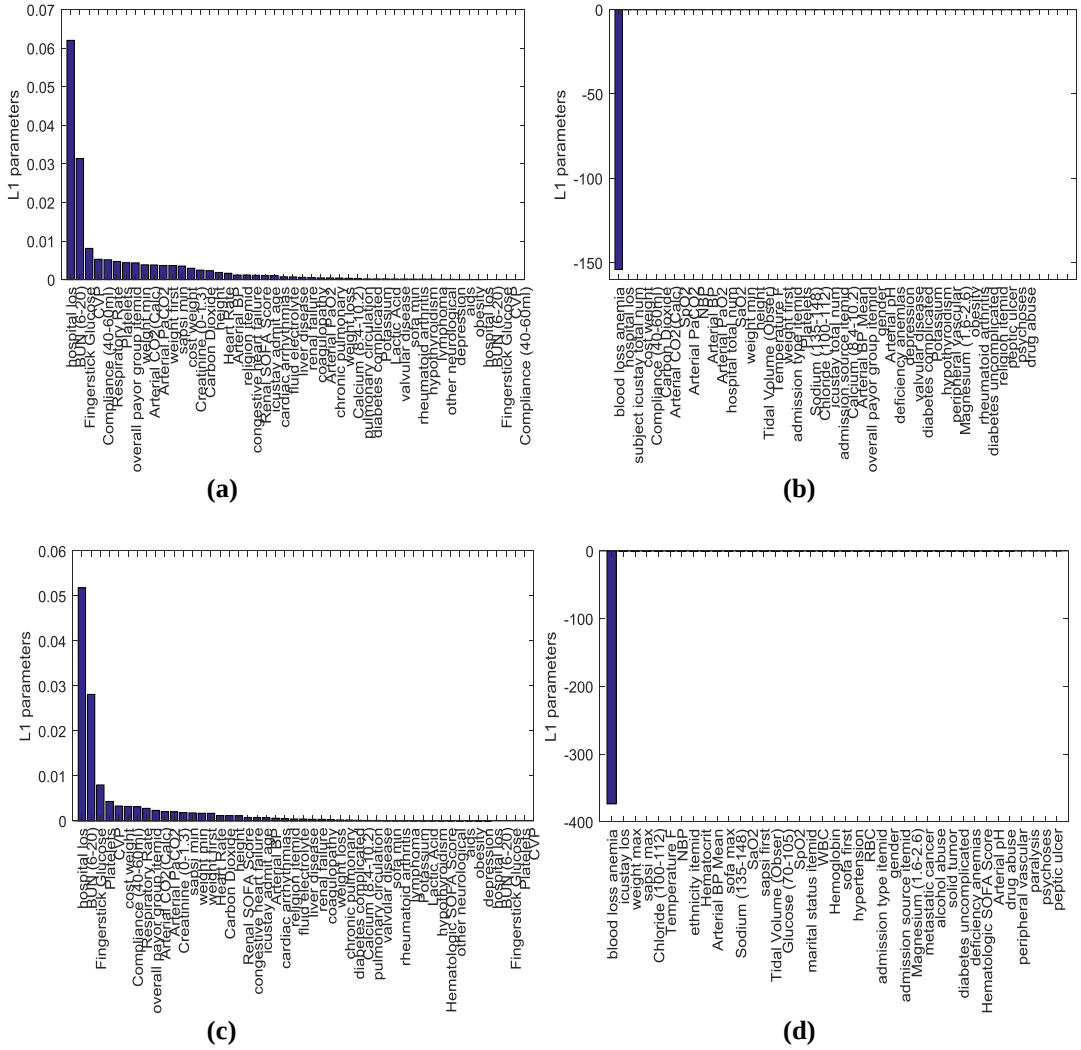


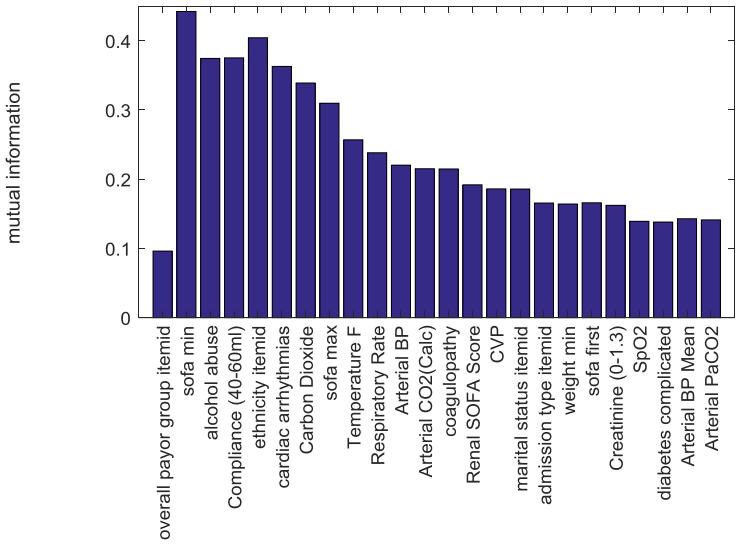
Figure 3.6: 30 day ICU readmission results showing the top features from LR with L1 regularization

(a) Gives a plot with the L1 regularized parameters which are correlated with ICU mortality (i.e. parameter value greater than 0 for the kmeans MNAR imputation (Imp1)

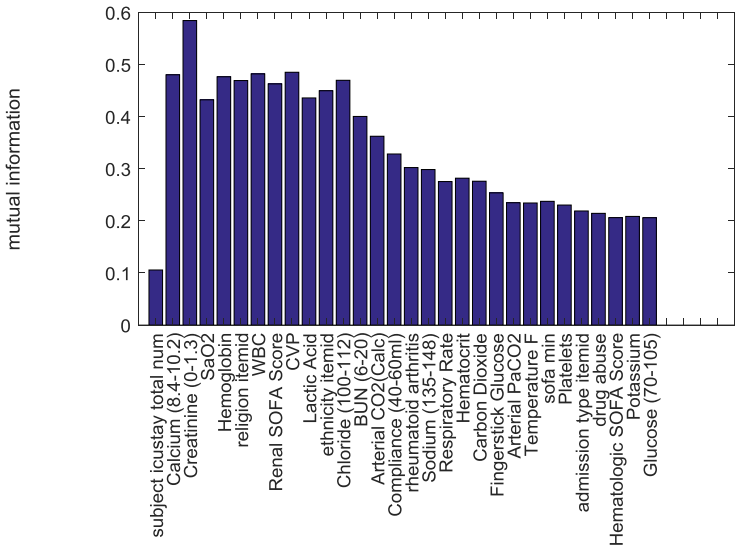
(b) Gives a plot with the L1 regularized parameters which are correlated with no ICU mortality (i.e. parameter value less than 0) for the kmeans MNAR imputation (Imp1)

(c) Gives a plot with the L1 regularized parameters which are correlated with ICU mortality (i.e. parameter value greater than 0 for the fcm MNAR imputation (Imp2)

(d) Gives a plot with the L1 regularized parameters which are correlated with no ICU mortality (i.e. parameter value less than 0) for the fcm MNAR imputation (Imp2)



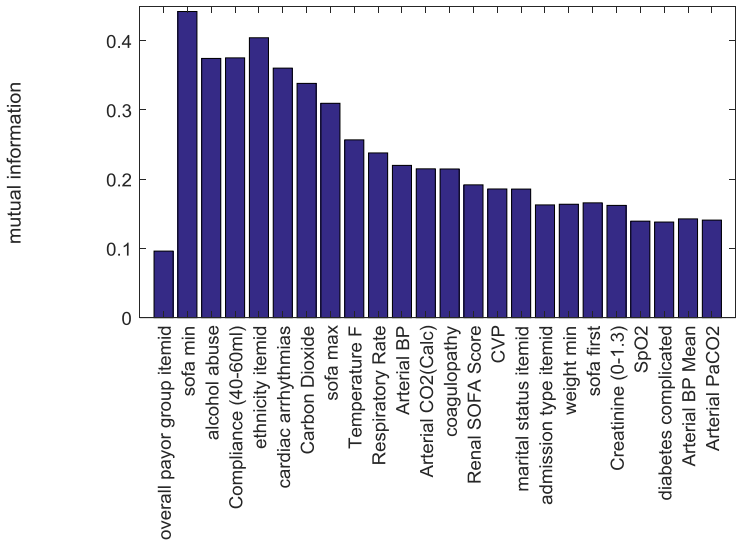
(a)



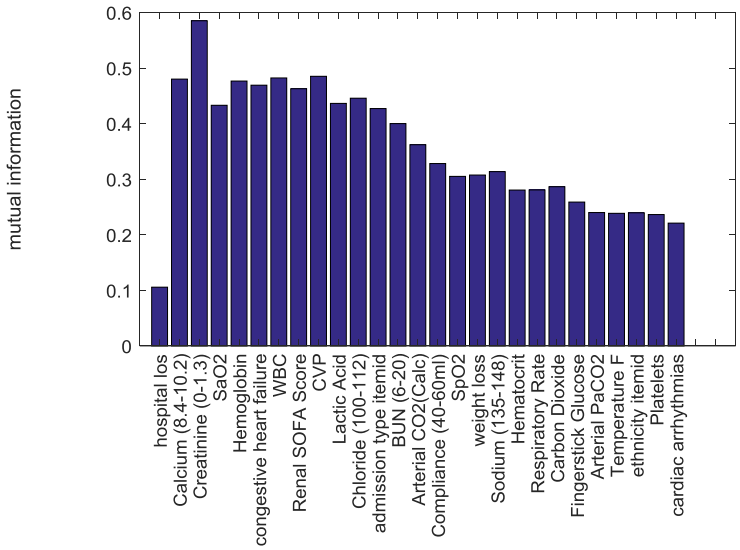
(b)

Figure 3.7: 30 day ICU readmission results showing the top features from mRMR with LR-
(a) Gives a plot of mutual information with kmeans MNAR imputation i.e. Imp-1
(b) Gives a plot of mutual information with fcm MNAR imputation i.e. Imp-2

Top ranking features predicted using our model (CRF) such as BUN; creatinine; cardiac and pulmonary disorders; ventilation; abnormal vital signs such as hypotension, hypertension, heart rate abnormalities, pO2 abnormalities, abnormal white blood cells



(a)



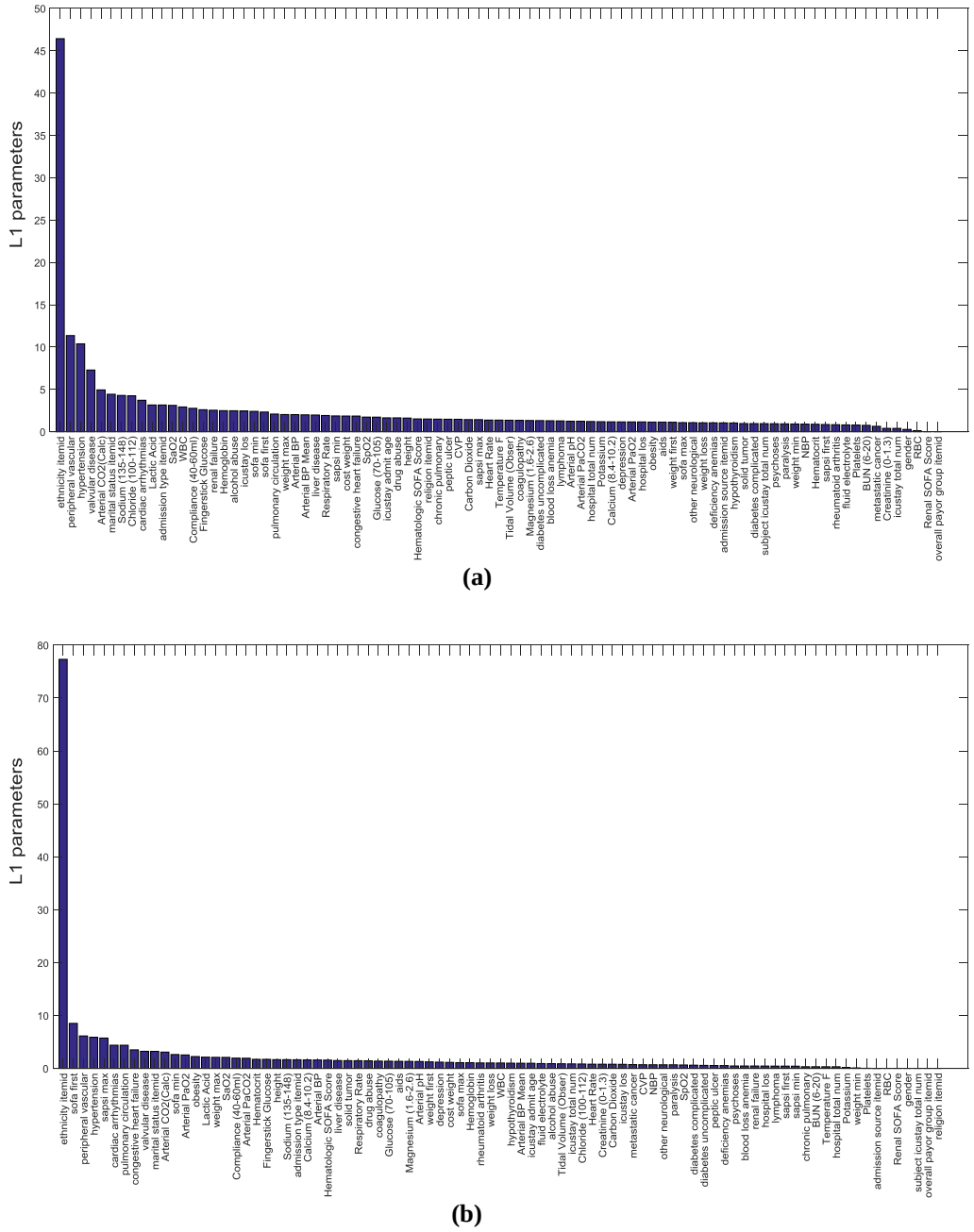
(b)

Figure 3.8: 30 day ICU readmission results showing the top features from mRMR with NN-
(a) Gives a plot of mutual information with kmeans MNAR imputation i.e. Imp-1
(b) Gives a plot of mutual information with fcm MNAR imputation i.e. Imp-2

(WBC) counts; abnormalities in lab results such as potassium, sodium, albumin; gastrointestinal and neurological symptoms have been clinically shown to be correlated with ICU readmission [185-190]. Outcomes such as sepsis, longer ICU length of stay and mortality have also been shown to be associated with ICU readmission [185].

Results from mortality analysis (Figure 3.9, Table 3.4) indicate that for CRF with L1 regularization, for Imp-1, physiological features and presence or absence of disease contributed most to decision. A total of 60 features contributed to 90% of the decision. Similarly Imp-2 physiological features and presence or absence of disease contributed most to decision with 45 features contributing to 90% of the decision.

For LR with L1 regression (Figure 3.10, Table 3.4), for both Imp-1 and Imp-2, the top features which correlated with mortality were the Simplified Acute Physiology Score (I) (SAPS-I) and Sequential Organ Failure Assessment (SOFA) Scores. The top feature which correlated with lack of mortality was the presence or absence of blood loss anemia. When the absolute value of the parameter values were used for computing the 90% contribution, the presence or absence of blood loss anemia contributed towards 90% of the distinguishing capacity between the two groups. When only the features which correlated with the presence of mortality was considered, a total of 12 features contributed to 90% of the classification. LR with mRMR (Figure 3.11, Table 3.4) the top features which contributed to classification were the Simplified Acute Physiology Score (I) (SAPS-I) and Sequential Organ Failure Assessment (SOFA) Scores, followed by physiological parameters. For Imp-1, 20 features contributed to 90% of the classification and for Imp-2, 23 features contributed to 90% of the classification. For NN with mRMR (Figure 3.12, Table 3.4) the top features which contributed to classification were the Simplified Acute Physiology Score (I) (SAPS-I) and Sequential Organ Failure Assessment (SOFA) Scores,



(a)

(b)

Figure 3.9: ICU mortality results showing the top features from CRF with L1 regularization-
(a) Gives a plot of mutual information with kmeans MNAR imputation i.e. Imp-1
(b) Gives a plot of mutual information with fcm MNAR imputation i.e. Imp-2

followed by physiological parameters. For Imp-1, 20 features contributed to 90% of the

classification and for Imp-2, 20 features contributed to 90% of the classification.

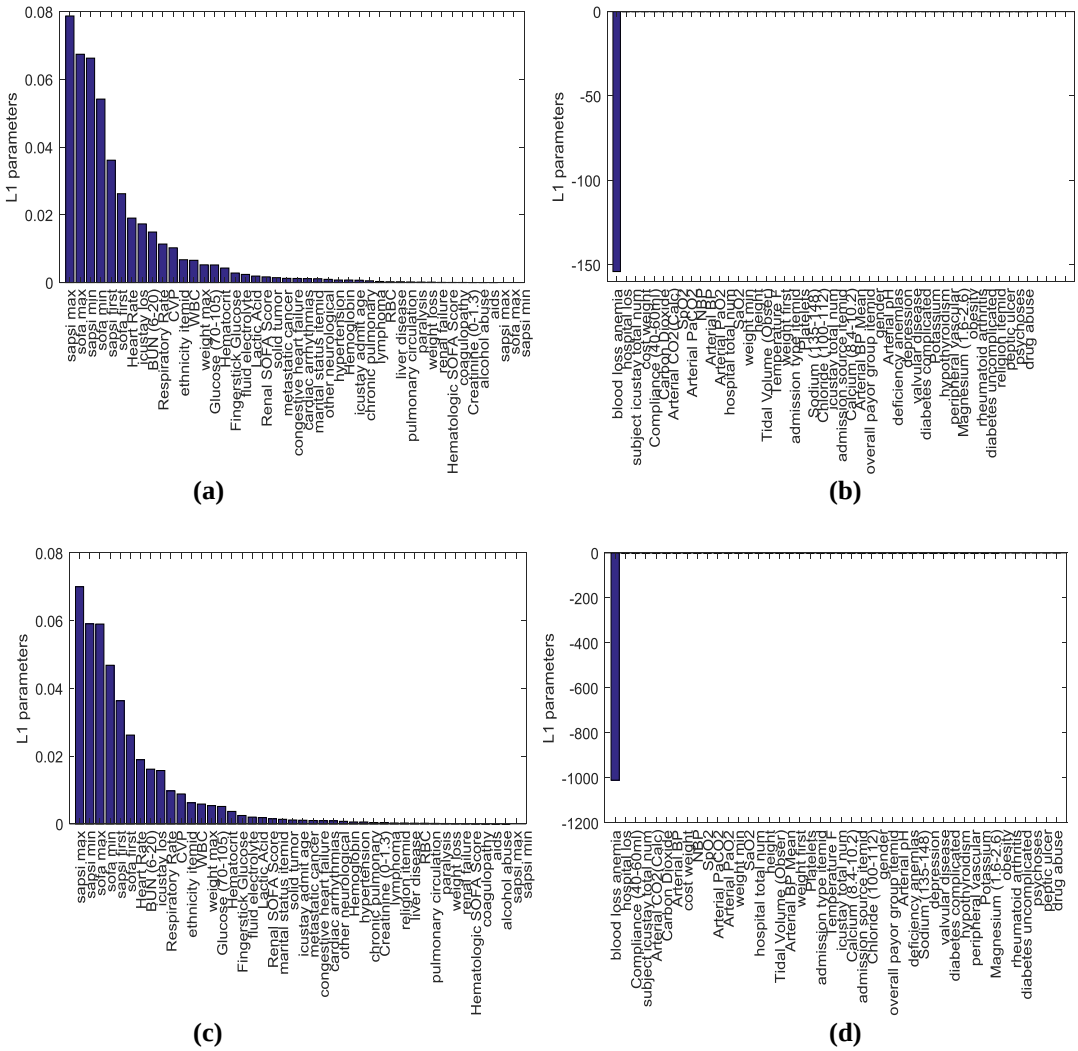


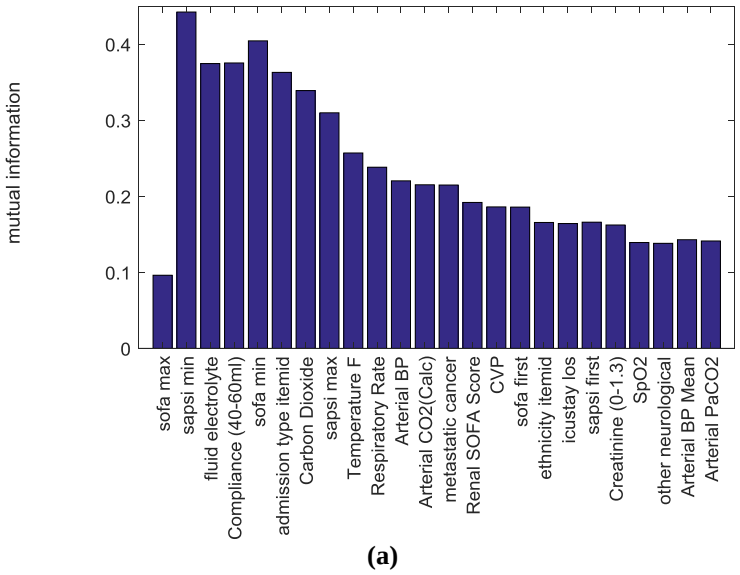
Figure 3.10: ICU mortality results showing the top features from LR with L1 regularization

(a) Gives a plot with the L1 regularized parameters which are correlated with ICU mortality (i.e. parameter value greater than 0 for the kmeans MNAR imputation (Imp1)

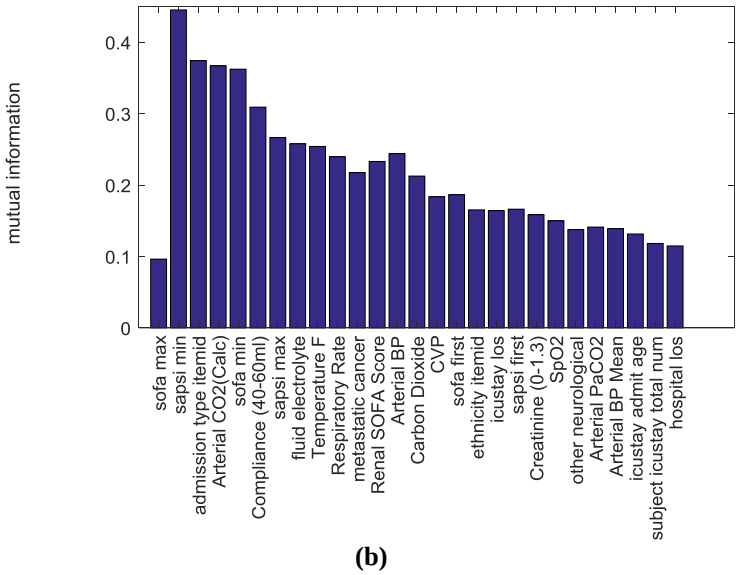
(b) Gives a plot with the L1 regularized parameters which are correlated with no ICU mortality (i.e. parameter value less than 0) for the kmeans MNAR imputation (Imp1)

(c) Gives a plot with the L1 regularized parameters which are correlated with ICU mortality (i.e. parameter value greater than 0 for the fcm MNAR imputation (Imp2)

(d) Gives a plot with the L1 regularized parameters which are correlated with no ICU mortality (i.e. parameter value less than 0) for the fcm MNAR imputation (Imp2)



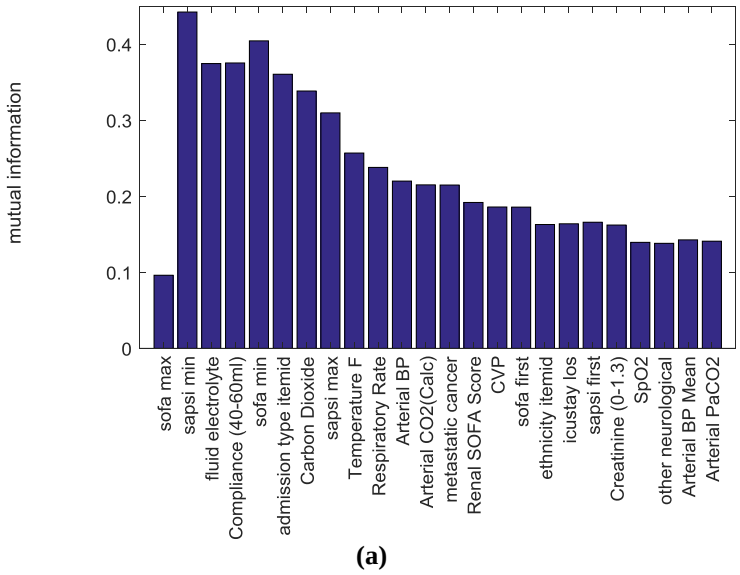
(a)



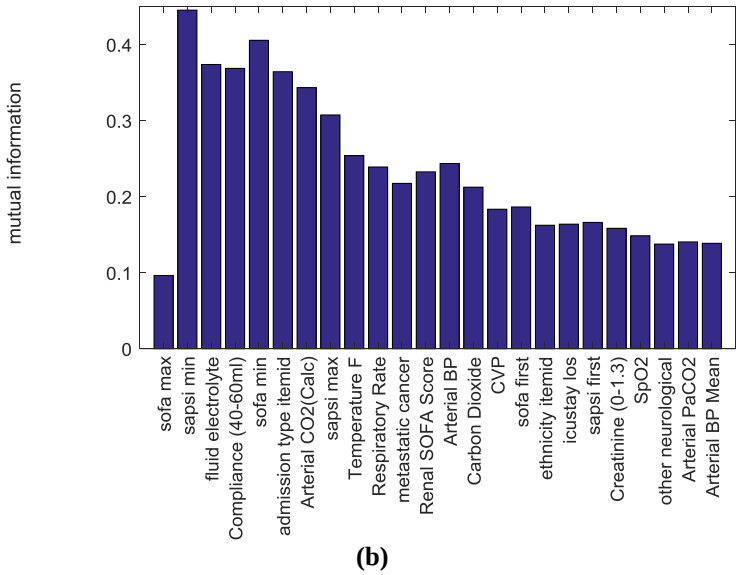
(b)

Figure 3.11: ICU mortality results showing the top features from mRMR with LR-
(a) Gives a plot of mutual information with kmeans MNAR imputation i.e. Imp-1
(b) Gives a plot of mutual information with fcm MNAR imputation i.e. Imp-2

Though the CRF models are not the best performing models, the features are different.



(a)



(b)

Figure 3.12: ICU mortality results showing the top features from mRMR with NN-

(a) Gives a plot of mutual information with kmeans MNAR imputation i.e. Imp-1

(b) Gives a plot of mutual information with fcm MNAR imputation i.e. Imp-2

Table 3.4:Top 5 features in LR, NN and CRF for all end-points. The list of all features with the contribution of each feature to the final decision are given in the appendix.

	Preprocessing	**LR with L_1**	**LR with mRMR**	**NN with mRMR**	**CRF with L_1**
ICU Readmission	Imp-1 (MNAR kmeans)	1.blood loss anemia 2. hospital los 3. BUN (6-20) 4. icustay los 5. weight max	1.overall payor group 2.sofa min 3.alcohol abuse 4.Compliance (40-60ml) 5.ethnicity	1.overall payor group 2.sofa min 3.alcohol abuse 4.Compliance (40-60ml) 5.ethnicity	1.Lactic Acid 2.weight max 3.metastatic cancer 4. obesity 5. cost weight
	Imp-2 (MNAR fcm)	1.blood loss anemia 2.hospital los 3.BUN (6-20) 4.icustay los 5.Fingerstick Glucose	1.hospital los 2.Calcium (8.4-10.2) 3.Creatinine (0-1.3) 4.SaO2 5.Hemoglobin	1.hospital los 2.Calcium (8.4-10.2) 3.Creatinine (0-1.3) 4.SaO2 5.Hemoglobin	1. sapsi max 2. hypertension 3. pulmonary circulation 4. sofa first 5. admission source\
ICU Mortality	Imp-1 (MNAR kmeans)	1.blood loss anemia 2.sapsi max 3.sofa max 4.sapsi min 5.sofa min	1.sofa max 2.sapsi min 3.fluid electrolyte amount 4.Compliance (40-60ml) 5.sofa min	1.sofa max 2.sapsi min 3.fluid electrolyte amount 4.Compliance (40-60ml) 5.sofa min	1.ethnicity 2.peripheral vascular disease (Y/N) 3.hypertension (Y/N) 4. valvular disease (Y/N) 5. Arterial CO2(Calc)
	Imp-2 (MNAR fcm)	1.blood loss anemia 2.sapsi max 3.sapsi min 4.sofa max 5.sofa min	1.sofa max 2.sapsi min 3.admission type 4Arterial CO2(Calc) 5.sofa min	1.sofa max 2.sapsi min 3.admission type 4Arterial CO2(Calc) 5.sofa min	1.ethnicity 2.sofa first 3.peripheral vasculardisease (Y/N) 4.hypertension 5.sapsi max

Top ranking features predicted using our model (CRF) such as SAP scores, long length of ICU stay, SpO2, comorbidities and SOFA scores have been clinically shown to be correlated with mortality [149, 152-158]. The features such as SAPS-I [149], ABP [159], age, heart rate, systolic blood pressure, body temperature, Glasgow Coma Scale,

mechanical ventilation, PaO2, FiO2, urine output, BUN (blood urea nitrogen), blood sodium, potassium, bicarbonates, bilirubin, white blood cells, chronic disease (AIDS, metastatic cancer, hematologic malignancy) and type of admission (elective surgery, medical, unscheduled surgery)[160] have been shown to be associated with mortality from other studies using the MIMIC-II dataset. In addition, SAPS-I, SpO2, creatinine have been shown to be associated with mortality in sepsis patients [157].

3.3.5 Visualization Results and Extension of CRF with Survival Analysis

Once prediction was obtained using CRF models, we developed an interactive graphical user interface (GUI), where the temporal risk profile for each patient could be viewed for each of the different end-points (Figure 3.6-3.7). The long-term goal for such a GUI is to assist physicians in adjusting treatment.

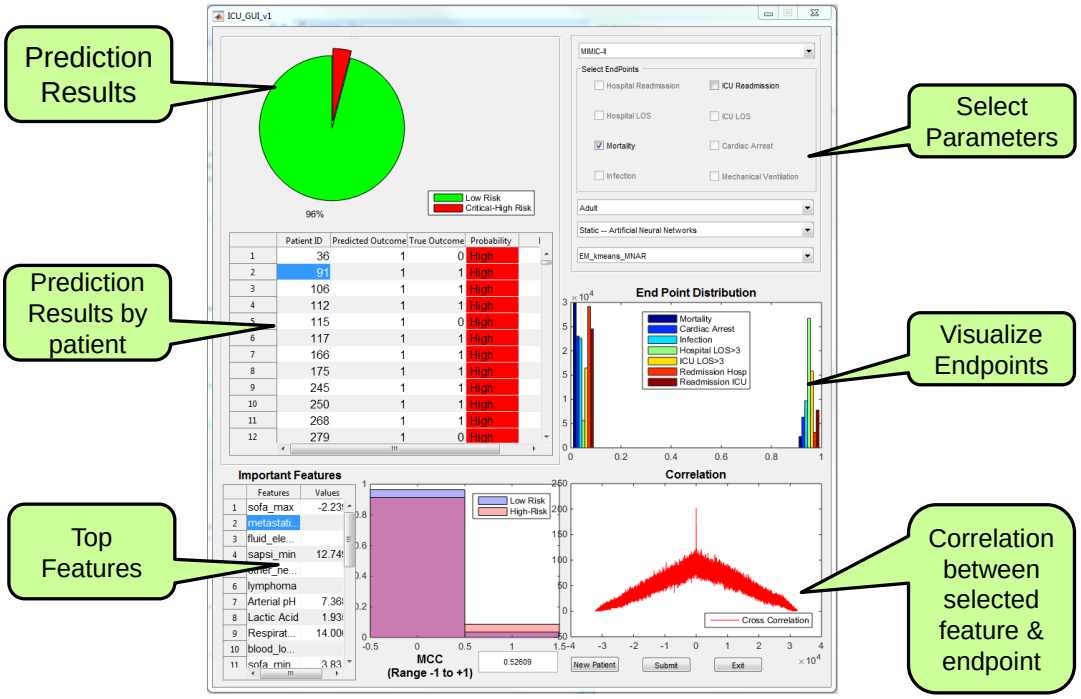


Figure 3.13: Interactive GUI and visualization for patient risk profiles.

Interactive GUI was developed for displaying the summary statistics of the population as a whole (Figure 3.6). The user can choose the dataset, end point, the preprocessing technique and the population of interest. On selecting this, the distribution of the age group of the population and the end-point distribution found in the population of interest are shown. In addition, the results of the model along with the probability of achieving the decision are seen in a graphical and a tabular format. The top features identified by the model along with the distribution in the dataset are also displayed. We also demonstrate the functionality to show the feature value for each patient along with the distributions for normal patients, critical patients with the actual value of the feature marked for each patient. We also allow real time input into the models

Our GUI (Figure 3.7) has the features to allow the physician plot the risk profiles, survival plots and also browse the top features and look at the temporal changes, changes in the rate and percentage changes in the different features used. Figure 3.7 shows the survival curve for a patient who is at high risk for ICU-mortality. We can see that as the

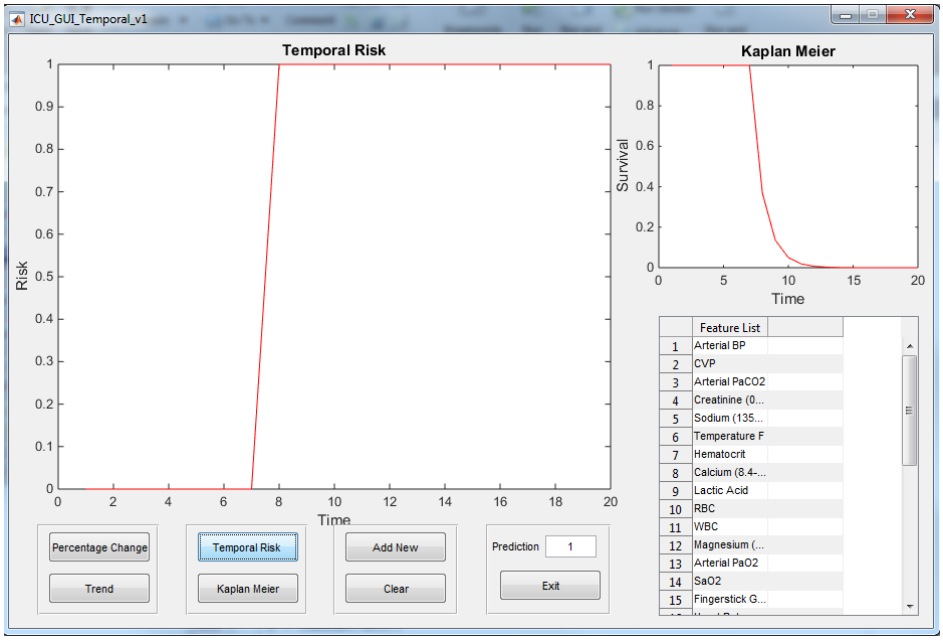


Figure 3.14: Temporal risk and survival curve.

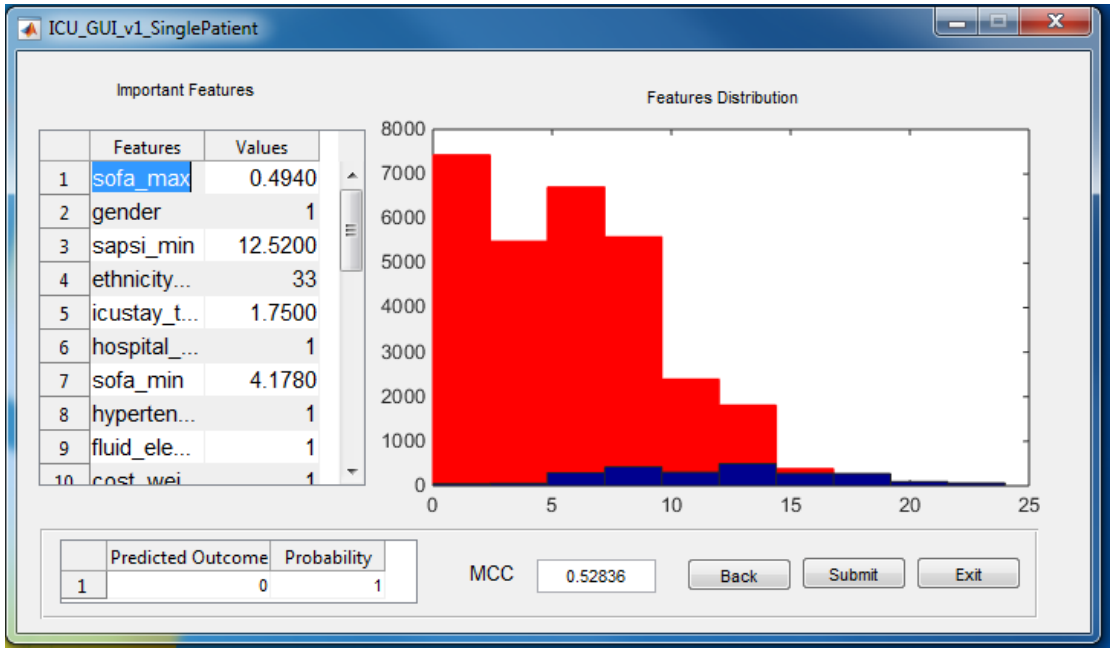


Figure 3.15: Single – User Screen – Non-Temporal.

risk increases, there is a corresponding drop in the survival curve, as evidence by the sharp downward slope. Such a temporal analysis is highly useful in the clinical setting and has a high potential for future clinical decision support.

In addition to this we have also developed New user screens (Figure 3.8) are developed for using the models to get the prediction for new patients. This screen is different for temporal and non-temporal analysis. The non-temporal analysis screen asks the values for the top features as selected by the model, displays the value with respect to the distributions for that feature and makes a prediction along with the probability. In addition, the MCC of the model is given to show the confidence which can be placed on this model.

3.4. Conclusion and Key Innovations

Prediction models for end-points such as ICU readmission and mortality remain challenging with limited efficacy in a wide variety of patients. State of the art ICU scores

do not include ongoing pathologic (acute) processes, processes which can impact long term outcomes as well. Our model sought to address this deficiency by looking at the interaction of these salient parameters for multiple end-points. In addition, the temporal nature of health records is not widely used for classification in current analytic models. Also, the current temporal models are not capable of providing a temporal risk profile for individual patients. We address these issues by utilizing a CRF based algorithm to find factors indicative of mortality in the ICU, and ICU readmission, using retrospective patient data. Our model was evaluated using 3×3 cross validation, where CRF outperformed most LR and NN models. It can be easily used to classify new patient data in an additive fashion without any retraining, hence proving scalability. As we had mentioned above, CRF models pick different features from LR and NN models and could essentially use different information to arrive at prediction results.

To summarize, the key innovations of this chapter include:

- Time series data analysis of ICU data without explicit independence assumptions
- Analysis of adults ICU data for mortality, and readmission
- First study to combine CRF with survival curves to show individual patient risk profiles

CHAPTER IV

COMBINATION OF STATIC AND TEMPORAL DATA ANALYSIS TO PREDICT MORTALITY AND READMISSION IN THE INTENSIVE CARE

4.1. Introduction

The modern intensive care unit (ICU) is a costly component of the national health care budgets accounting for 13.7% of hospital costs and 4.1% of national health expenditures [191-194]. These costs are largely explained by adverse outcomes such as prolonged length of stay in the ICU and ICU readmissions [195, 196]. For these reasons, there has been substantial research in developing clinical decision support systems to predict and prevent ICU outcomes, including ICU mortality, and ICU readmission.

Current research on the use of critical care data has focused on the use of either static data (these are generally fixed variables like gender, socioeconomic status, weight on admission) , temporal data (such as heart rate, blood pressure, lab tests) or continuous data (such as ECG, ECG). Conventional static data analysis using methods such as Cox regression and logistic regression though very useful for finding risk factors associated with a specific disease, do not incorporate the temporal nature of the clinical data. Similarly, temporal models such as sequence analysis and association rule mining [85, 87, 88] and temporal Cox regression [89-91] generate models using the temporal nature of data. However, most of the current work suffer from challenges such as the lack of data

analytics that can make sense of patient conditions using a combination of static and temporal data (sequential and continuous).

In the previous chapter, we performed a temporal analysis using conditional random fields (CRF) to predict ICU mortality and 30 day ICU readmissions using adult patient data from a publicly available database called MIMIC II [197]. We compared our methods using conventional analysis of logistic regression (LR) and neural networks (NN). From our analysis we found that more temporal features were selected by CRF models and included features such as arterial BP, central venous pressure, creatinine, arterial PaCO2. In contrast, the LR and NN models picked features such as max sequential organ failure assessment (SOFA) score, metastatic cancer, minimum simplified acute physiology score (SAPS) I and presence of neurological symptoms. In addition, the data in the ICU itself is collected at higher sampling rates, though this can also vary.

In this study we extend our previous work to demonstrate a framework with which we can combine data from multiple sources, sampled at different sampling frequencies (e.g. static and temporal models (sampled at 6 hour intervals)) using ensemble techniques such as hard and soft voting. The static models include logistic regression and feed-forward neural networks, and the temporal models include conditional random fields. We combined the decisions from these individual classifiers and demonstrate our results using adult data from Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC) –II.

4.2. Methods

In this work, we perform a retrospective analysis of ICU data for adult patients to demonstrate the advantages of the combination of static and temporal data mining. After data preprocessing, we perform static data analysis using logistic regression and feed-

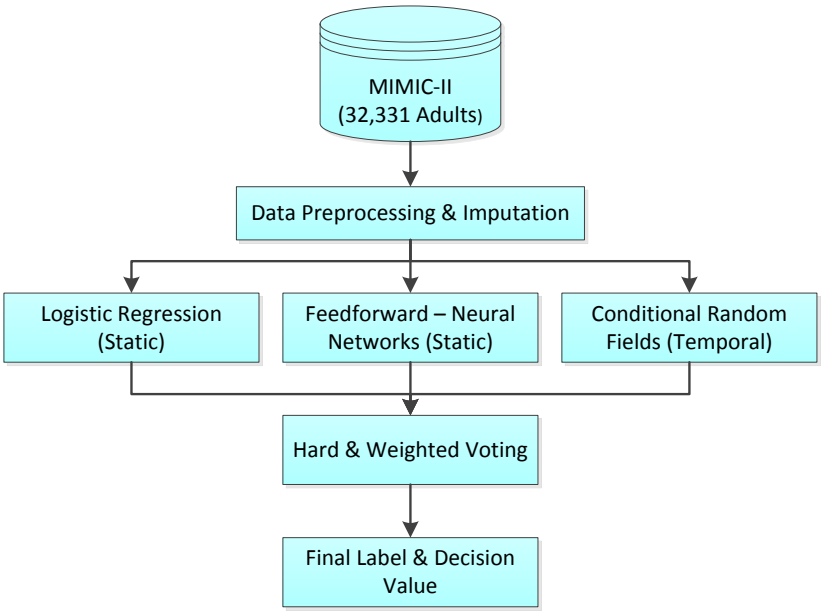


Figure 4.1: Combining static and temporal models.

forward neural networks, and temporal data analysis using conditional random fields. We then combine the decisions of these different classifiers using hard and soft voting techniques (Figure 4.1.).

4.2.1 Data Preprocessing

The pre-processing of data for non-temporal analysis was performed by averaging the temporal data over the duration of stay. For temporal analysis, we binned the data to reduce the effects of missing data. Then outliers whose values were physiologically impossible were removed. If the value is normally distributed, then values that deviated by ±3 standard deviations from the mean value were also removed. The missing data was divided into the three types mentioned in Chapter 1 ("Neglectable", Recoverable" and "NER"). Then each type was imputed differently using the techniques described in Chapter 1 [180]. "NER" data was imputed using student's t-copulas and "Recoverable" data was imputed using expectation maximization (EM) after clustering [180]. Both kmeans

("NER" kmeans) and fuzzy C means ("NER" fcm) were used for clustering the data prior to imputation here. We will refer to these two imputation techniques as 'Imp-1' and 'Imp-2'.

4.2.2 Data Mining on Static Data

For the analysis of static data, we use logistic regression and feed-forward neural networks, which are the most commonly used models in healthcare, to predict the patient outcomes of the study, ICU mortality and 30 day ICU readmission. A logistic regression model is trained for each of the outcomes using a feature set $X = \{x_1, x_2 \ldots x_n\}$ derived from the clinical measures mentioned above. Logistic regression model calculates the probability of adverse ICU outcome given by (4.1)

$$h_\theta = \frac{e^{\theta 0 + \theta 1 x1 + \theta 2 x2 + \cdots \theta nxn}}{1 + e^{\theta 0 + \theta 1 x1 + \theta 2 x2 + \cdots \theta nxn}} \quad (4.1)$$

The outcome group (y) is assumed to be true (1) when the probability h_θ exceeds a certain threshold. The values of parameters $\theta = (\theta_0, \theta_1, \theta_2, \ldots \theta_n)$ are trained from the training data set by maximizing log-likelihood. In order to prevent over fitting we used L_2 regularization and minimum-redundancy maximum-relevancy (mRMR) for feature selection [198]. Hence the hyper parameters to be trained include the regularization parameter and the number of features.

Feedforward neural networks (ANN) are essentially mathematical models defining a function $f : X \rightarrow Y$ or a distribution over input (X) or both input (X) and outcome (Y). The neural network consists of many interconnected nodes with each input from the input layer being fed up to each node in the hidden layer, and from there to each node on the output layer. The hyper-parameters of the model include the number of nodes and layers when optimizing the neural network. In this study, the number of input layer nodes equaled

the number of features from which an optimal number was selected using mRMR and the number of hidden layers equaled 1. Hence, the hyper parameters optimized were the number of hidden layer units and the number of features selected using mRMR. The optimization of the hyper parameters for both these techniques were performed using 3×3 nested cross-validation.

4.2.3 Data Mining on Temporal Data

For the analysis of temporal patient data we used conditional random fields (CRF) [199]. CRF represents the conditional probability of the outcome, $y \in Y$ given a sequence of ICU measurements $x = \{x_1, x_2 \ldots x_T\}$ i.e. $p(y|x, \theta)$, where θ is the set of parameters. In addition we also assume certain hidden variables $h = \{h_1, h_2 \ldots h_m\}$ derived from the combination of features at each time point. The hidden states h take a value from a finite set of values given in H. The probability $P(y, h|X, \theta)$ is given by (4.2).

$$\mathbf{P(y, h|X, \theta) = \frac{1}{Z} e^{(\theta \varphi(y,h,x;\theta))}} \tag{4.2}$$

where θ is the set of parameters estimated during training, $\varphi(y, h, x; \theta)$ is the clique potential function, and a clique is a fully connected sub-graph [183]. Cliques in a chain CRF (used here) consists of an edge between adjacent labels (y_{t-1} and y_t) as well as the edges from those two labels to the set of observations x. As a result, CRFs represent the conditional probability as (4.3-4.6):

$$P(y|x, \theta) = \sum_h \frac{1}{Z} e^{(\theta \cdot \varphi(y,h,x;\theta))} \tag{4.3}$$

where,

$$Z = \sum_{y,h} e^{(\theta \cdot \varphi(y,h,x;\theta))} \tag{4.4}$$

$$\varphi(X, h, Y; \theta) = \sum_{j=1}^{T} \sum_{l \in F} f_l^1(j, y, h_j, X) \theta_l^1 + \sum_{j,k \in E} \sum_{l \in F} f_l^2(j, k, y, h_j, h_k, X) \theta_l^2 \tag{4.5}$$

where, E, F are the number of edges and features respectively. And f_l^1, f_l^2 are feature transformation functions (analogous to regression here). Hence, the likelihood function is given by equation 4.3

$$P(Y|X) = \frac{1}{Z(X)} \times \prod_{i=1}^{n} exp(\sum_h \varphi(x_i, h, y_i; \theta)) \quad (4.6)$$

The log-likelihood is maximized to learn the parameters θ. The inference is done by forward-backward inference to obtain the outcome probability from the graph. Over-fitting of the CRF model is prevented by L_1 regularization of weights (the absolute values of weights are penalized). The optimization of the hyper parameters such as the number of hidden states and the L_1 regularization coefficient was performed with 3×3 nested cross-validation [147].

4.2.4 Combining Static & Temporal Models using Hard & Weighted Voting

The decision values and decisions from the 3 classifiers were combined using hard and weighted voting techniques. We tested a total of four different methods to combine the decision or decision values. In the first method (M1), combined the three classifiers by hard voting where the majority value of the decision (mode of the three decisions) was used as the label. In the second method (M2), we used the mean of the decision values from the three classifiers to get a new decision values which was used to compute the label. The next two methods involved weighted voting, where we first weighted the decisions. The weights for each classifier was computed as follows (4.7)

$$Weight = \log\left(\frac{Cl_{Per}}{1-Cl_{Per}}\right) \quad (4.7)$$

where Cl_{Per} is the classifier performance (Matthews Correlation Coefficient (MCC) scaled between 0 and 1). The decision values (M3) was computed as a weighted average of the decisions. This decision value was used to obtain the final label. In the last

method (M4), the weights for each classifier was obtained using (8). The final decision value was the weighted average of the individual classifier decision values. The computed decision value was then used to compute the final label.

4.2.5 Evaluation of the Classification Methods

The evaluation of all the combination methods was performed using 10-fold cross validation. We repeated the process 3 times and report averaged values of Matthews correlation coefficient (MCC) and accuracy. We chose MCC as a metric because of its relative tolerance to an imbalanced population.

4.3. Results

4.3.1 Case Study: Adult ICU Database

Data Source – MIMIC-II Database

Table 4.1: Top 5 Feature Types in Dataset.

Data Type	**Examples of Measures**
Demographics	Gender, Age, Height, Weight, Ethnicity, Comorbidity
Lab Data	Urea, Albumin, Bilirubin, Creatinine, Sodium
Chart Data	HR, BP, Arterial PH, Arterial $PaCO_2$, Arterial PaO_2

This study is a retrospective data analysis using data from Multi-parameter Intelligent Monitoring in Intensive Care, second version, (MIMIC-II) database. MIMIC-II is a public ICU data repository with 32,331 adult and 8,080 neonatal records [149], mentioned in the previous chapters. The features included physiological measures (e.g. heart rate, blood pressure), lab results (e.g. while blood cells, red blood cells, cholesterol), administrative data (e.g. length of stay), diagnostic codes (ICD-9), and comorbidities (Table 4.1).

This dataset contains 2,334 patient records with mortality during the ICU stay and 29,997 patient records of successful discharge from the ICU. Similarly, 7,787 patient records had an ICU readmission within 30 days and 24,544 patients did not relapse into the ICU within 30 days. As mentioned above, we first performed classification using static and temporal classification methods and then combined the decision values and decisions using voting methods.

Results for Adult ICU Data to Predict ICU Mortality and ICU-Readmission

The results from individual classifiers and the combined models are shown in Tables 4.2 and 4.3. Our results indicate that the combination models outperformed the individual models when using both MCC and accuracy as the metrics for the endpoint of mortality. The methods of combining decision values and weighted voting methods have the best MCC. The best performing combination models give an improvement in MCC of 6-7% over logistic regression, 2% over neural networks and 3-8% over conditional random fields for mortality. For 30 day ICU readmission, all the combination models performed better than the static models for both imputation techniques used. For Imp2, the temporal models performed better than the combination models. When MCC was used as the metric for comparison, the methods of combining decision values and weighted voting methods gave the best performance. The best performing combination models give an improvement

Table 4.4: Classification Results from ICU Mortality (Mathews Correlation Coefficient) (LR = Logistic regression, NN = Neural networks, CRF = Conditional random fields, M1 = Voting, M2 = Mean of decision values, M3 = Weighted mean of decisions, M4 = Weighted mean of decision values).

	Imputation	LR	NN	CRF	M1	M2	M3	M4
Mortality	**Imp1**	0.47 ± 0.006	0.52 ± 0.006	0.51 ± 0.033	0.52 ± 0.003	0.54 ± 0.004	**0.54 ± 0.006**	**0.54 ±0.006**
	Imp2	0.48 ± 0.009	0.52 ± 0.007	0.46 ± 0.098	0.52 ± 0.002	0.54 ± 0.005	**0.54 ± 0.005**	**0.54 ± 0.006**
Readmission	**Imp1**	0.32 ± 0.005	0.39 ± 0.004	0.39 ± 0.021	0.59 ± 0.039	0.65 ± 0.001	**0.65 ± 0.001**	**0.65 ± 0.001**
	Imp2	0.33 ± 0.007	0.39 ± 0.003	**0.73 ± 0.032**	0.58 ± 0.031	0.66 ± 0.003	0.66 ± 0.002	0.66 ± 0.003

Table 4.3: Classification Results from ICU Readmission (Accuracy) (LR = Logistic regression, NN = Neural networks, CRF = Conditional random fields, M1 = Voting, M2 = Mean of decision values, M3 = Weighted mean of decisions, M4 = Weighted mean of decision values)

	Imputation	LR	NN	CRF	M1	M2	M3	M4
Mortality	**Imp1**	0.94 ± 0.001	0.95 ± 0.000	0.95 ± 0.003	**0.95 ± 0.000**	**0.95 ± 0.000**	**0.95 ± 0.001**	**0.95 ± 0.001**
	Imp2	0.94 ± 0.001	0.95 ± 0.001	0.94 ± 0.004	**0.95 ± 0.000**	**0.95 ± 0.000**	**0.95 ± 0.001**	**0.95 ± 0.001**
Readmission	**Imp1**	0.79 ± 0.001	0.80 ± 0.000	0.80 ± 0.006	0.86 ± 0.013	0.86 ± 0.000	**0.87 ± 0.000**	0.86 ± 0.000
	Imp2	0.79 ± 0.001	0.80 ± 0.001	**0.90 ± 0.013**	0.85 ± 0.010	0.87 ± 0.001	0.87 ± 0.001	0.87 ± 0.001

in MCC of 33% over logistic regression, 25-26% over neural networks and 26% over conditional random fields for Imp1. The readmission models with Imp-2 performed better than combination models for ICU readmission.

4.4. Conclusion and Key Innovations

Prediction models for clinically significant end-points such as ICU readmission remain challenging with limited efficacy in a wide variety of patients. In addition, ICUs also collect data at different frequency rates. In this work, we combine static models, such as logistic regression and feedforward neural networks, with temporal models such as conditional random fields (CRF), by hard and weighted voting techniques. The combined models gave a better performance as compared to individual models. The weighted models

where the proportion of the decision making was based on individual performances gave the best overall performances. We can conclude that combination of multiple model types with different feature types improves the robustness of the model for complex data types and hence has the potential to enhance immediate management of a patient and the overall resource utilization.

Our work, currently combines data from only adult patients from MIMIC-II and also does not include high frequency data such as waveform data. In the future we aim to overcome these challenges and demonstrate our results on pediatric data from Children's Healthcare of Atlanta after IRB approval. We also aim to combine intermediate features using deep-learning approaches.

To summarize, the key innovations of this chapter include:

- Combination of static and temporal data for mortality and 30 day ICU readmission.
- We show an improvement for integration models over individual modalities.

CHAPTER V

DEEP MODELS FOR INTEGRATING TEMPORAL DATA WITH STATIC DATA TO PREDICT ICU LENGTH OF STAY IN CHILDREN

5.1. Introduction

The long length of stay (LOS) in the hospitals and intensive care unit (ICU) is a key contributing factor towards the higher cost of healthcare and is associated with long-term adverse effects such as neuro-developmental disorders for young kids [18]. Patients with an ICU stay longer than seven days utilize more than 50% of the ICU resources [200]. Studies have shown that factors such as intensivist consultation and admission standards [201], improved sedation practices, oxygen therapy for high-risk surgical patients[202], and effective communication [200] have been instrumental towards reducing ICU lengths of stay. Hence, the identification of patients at risk for a longer ICU stay is essential for the effective allocation of resources.

The ICU is a complex environment with large amounts of multimodal, multi-time resolution data collected for each patient visit. ICU data is collected in electronic health records (EHR) systems at multiple temporal resolutions and can be categorized as static/ non-temporal, temporal, and continuous waveform. Static data does not change over the patient stay. Temporal data such as clinical parameters, lab tests, and medication change several times during a single ICU stay. Continuous waveform data such as electrocardiogram, electromyogram, and electroencephalogram have multiple values each second. This data collected in EHR systems can be leveraged to predict clinical outcomes such as risk for long ICU stay.

Scores which help predict LOS using electronic health record (EHR) data (APACHE [203], and SAPS [204]) are primarily multivariate regression-based models. These models either use aggregated information or temporal data alone for making predictions about ICU outcomes such as length of stay. These models do not combine temporal and non-temporal data in the EHR. In this study, we propose the integration of temporal data (binned into 2-hour intervals) and non-temporal data using deep-learning methods. Temporal data from EHR is used for LOS prediction using long-short-term-memory networks (LSTM), and non-temporal data is modeled using neural networks. . We compared our models against standard classifiers such as k-nearest neighbors (kNN), random forests, decision trees, and support vector machines (SVM), and regression models such as linear, SVM, decision trees and random forest regression models We also demonstrate the superiority of combination models against the individual models. Finally, we focused on a pediatric population which tends to be much more heterogenous and under-studied.

Longer ICU length of stay is associated with higher risk of life-threatening outcomes such as long-term mortality [28, 29, 154], acute kidney injury [21, 29], sepsis and severe infections [23, 24]. Any intervention, pharmacologic or procedural, that could abbreviate their length of stay in ICU would have a significant impact on the child's quality of life and society's overall resource utilization [34-36].

Current research on ICU LOS use aggregated data and are based on standard machine learning techniques such as regression trees [205, 206], support vector machines [207, 208], random forests [206, 207], Bayesian methods [208], neural networks [209] and kNN [208]. The challenge with these models is that only the first-day data [210, 211] or

aggregated models can be used. These challenges are addressed by building temporal models such as Markov processes [212] and Cox regression [213]. However, these models do not integrate temporal and non-temporal data from LOS prediction.

The combination of data at different temporal resolutions (temporal and non-temporal (static) data) can be performed either at the feature level, the decision level or at intermediate feature level [214, 215]. The combination at feature level can be performed through aggregation or abstraction of the temporal data, or through repetition of static data [216, 217]. The combination at decision level can be performed through hard and soft voting principles [218]. However, both these types of combinations do not account for the temporality of data and the interaction between the different types of data. In this work, we showcase the combination of the two data types using intermediate features through a deep learning approach.

Deep learning models have shown great potential for prediction in EHR data [114, 120]. In addition to non-temporal models such as autoencoders [112, 114], deep belief networks [219], and restricted Boltzmann machines [117], temporal models such as recurrent neural networks (gated recurrent units (GRU) [120], and LSTMs [220]) have shown improvement over standard techniques. LSTM models [221], in particular, have proven effective in the areas of sequence mining such as natural language processing [222], handwriting recognition [223], speech recognition [224], and bioinformatics [225, 226]. In EHR data analysis, LSTM and GRU models have been used for the classification of disease using both waveform and time-series data with proven improvement over standard machine learning [120, 123]. LSTMs are particularly effective for EHR sequence analysis since they are capable of handling sequences of varying lengths, can handle long-term and short-term

dependencies, and are tolerant towards missing data. Since LSTM networks utilize temporal relationships to provide intermediate features, they can be integrated with the non-temporal model. Common EHR data analysis methods for non-temporal analysis include logistic regression [28, 33, 78], cox regression [78] and artificial neural networks [28, 80, 170]. Feed-forward neural networks have been widely used for classification and are amenable towards integration with LSTMS.

In our analysis, we integrate temporal and non-temporal data for the length of stay prediction using pediatric data. For temporal data, we use LSTM models which we integrate with neural networks for non-temporal data. We use the models for predicting patients at a high risk of LOS >& and for predicting the ICU LOS. We evaluate our classification models using metrics of accuracy, Matthews correlation coefficient (MCC), precision, recall and F1 scores, and regression models using root mean square error (RMSE) and mean absolute error (MAE). . The major contributions of this paper are as follows:

- We developed a framework for combining data from multiple temporal resolutions using deep models for classification and regression tasks.
- We show the effect of using long term associations on the prediction performance of LSTM models.
- We demonstrate ICU length of stay prediction on pediatric ICU patient data
- We developed perturbation based analysis for data interpretation.

We structure the rest of the chapter as follows: we first describe the modeling and evaluation framework; followed by the results and discussion. Here where show an interpretation of our models followed by the conclusions.

5.2. Methods

In this study, we perform a retrospective data analysis using pediatric patient records. Since the majority of the ICU resources (> 50%) are utilized by patients with LOS greater than seven days we classify the patients with stay longer than seven days from those whose stay is lower than seven days. (median ICU stay is 3.8 days) [200]. In addition, we also make predictions about the ICU length of stay using regression models.

5.2.1 Data Description

Our dataset is from Children's Healthcare of Atlanta (CHOA) containing 5000 patient records spanning an 11 month period. The visits spanned pediatric ICU (PIC), Neonatal ICU (NICU), and cardiac ICU. Each ICU stay record consists of the patient's demographic information (e.g., gender and age of admission), diagnosis (e.g., ICD-9 codes), birth-related events (e.g., birth weight, head circumference, gestation weeks), microbiology events (e.g., microbes in blood or serum), chart events (e.g., heart rate), medication intake events, microbiology events (e.g., microbes), and clinical records (e.g., heart rate, oxygenation) collected from bedside monitors, averaged over each min (Table 5.1.)

The data columns were binary, categorical and quantitative from which we extracted features. We used categorical data such as the disease codes and procedure codes into the number of times each disease or condition was present or the procedure performed.

Table 5.1: CHOA Data Description

Data Type	Examples of Measures
Demographics	DOB, Gender, Age, Height, Weight, Ethnicity, Religion, Date of Death, Co morbidity with other diseases
Microbiology	Types of microbes, Amount of microbes, dilution
Lab Data	Urea, Albumin, Bilirubin, Creatinine, Sodium, Potassium, Calcium
Medication Data	Medication & IV administered, Dosage, Duration time, Concentrations & Rate of Administration, composition of IV imposed

This gives us 9,071 non-temporal features consisting of demographics, microbiology, diagnosis codes, and medication data. In this dataset, since the temporal information for microbiology, medication, and pathology was not available, we treated them as non-temporal data and performed aggregates over the duration of the stay. The temporal data we used was from the various lab tests performed. Lab test data had a median sampling interval of 2.05 hours, from which we extracted 2,500 features. After removing features with greater than 80% missing data, we were left with 1,882 lab features, which we binned into 2 hours binning interval. In addition, this dataset had an issue where the tests or values were not recorded for very long time intervals (~ several days) in the middle. This could be due to the fact that the patients were no longer in the ICU. We treated this type of data as multiple time-series for each patient visit, and did not use the missing period for binning. Since non-temporal data has less than 1% data missing no features were removed.

5.2.2 Data Preprocessing and Feature Selection

The features were extracted in the previous step were either quantitative real numbers or binary. The range of quantitative features had an order of magnitude variation (e.g. respiration rate varied from 10 – 30 breaths per min, blood pressure varied from 90-150 mmHg, and blood calcium varied between 8-11 mg/dL). To address this issue, we normalized all the features between the ranges 1 – 2. We also converted the binary values into 1 or 2. Following this, we performed feature selection for temporal data using minimum redundancy maximum relevance (mRMR) [227]. We tested for 200, 300 and 400 features. For mRMR, we used mean interpolation, however in the subsequent analysis, no data imputation was used. In the future, we will investigate the use of more sophisticated feature selection techniques which can handle partial data, such as mutual information with

missing data [228] and margin based feature selection [229]. Following feature selection, we proceed to classification with individual classifiers (non-temporal and temporal) followed by the integration.

5.2.3 Length of Stay Prediction using Non-Temporal Data

We performed the analysis of non-temporal data (features from administrative data, demographics, diagnostic codes, microbiology, medication, and pathology) with 9,071 features using feedforward neural networks. Feedforward neural networks can be defined as a function which gives the relationship between input X and outcome Y (length of ICU stay >7). Each layer of a feedforward neural networks takes an input of dimension $n \times d$, where n is the number of samples d is the input dimensionality to give z (5.1)

$$z = s(Wx + b) \tag{5.1}$$

where s.is an activation function such as sigmoidal or tanh, $[W, b]$ are parameters to be trained. The inputs are passed through a number of layers till the final layer gives the classification output Y. The model parameters are trained using backpropagation on the training dataset. We used a total 3 hidden layers and tanh activation function. The size of each hidden layer was half of the previous layer. Hyper-parameters such as the network size was determined using a grid search. We tested a network size of 2 and 3 layers, with the size of each of the first hidden layer 400, 300, 200.

5.2.4 Length of Stay Prediction using Temporal Data

As mentioned above we selected 400 temporal features from lab data, which we use to classify the patients into LOS >7 days and those with LOS < 7 days. We used LSTM networks for temporal analysis. Several variations of LSTM networks have been proposed since their introduction by Hochreiter et. al. The most commonly accepted version of LSTM networks today [230, 231] is composed of units called memory blocks, where each memory block contains three gates (input, output and forget) and peepholes. Input gates are used for regulating (scaling) the input into the memory cells. Output gates are used for regulating the outputs to the rest of the network. Forget gates are used for adaptively resetting the states within memory cells. Peepholes allow three gate values from within each cell to modify the current state within each memory block. The regulation of each memory block (Figure 5.1.) is governed by equations 5.2 – 5.8, which are iteratively calculated at each time-step in the network. The input at each time instant t (x_t), is passed through the input layer to get the value at the input gate e_t. e_t is then passed through the other 2 gates and using the values at each peephole we get the output y_{t+1}.

$$y_{t+1} = softmax(W^{ym} m_t) \tag{5.2}$$

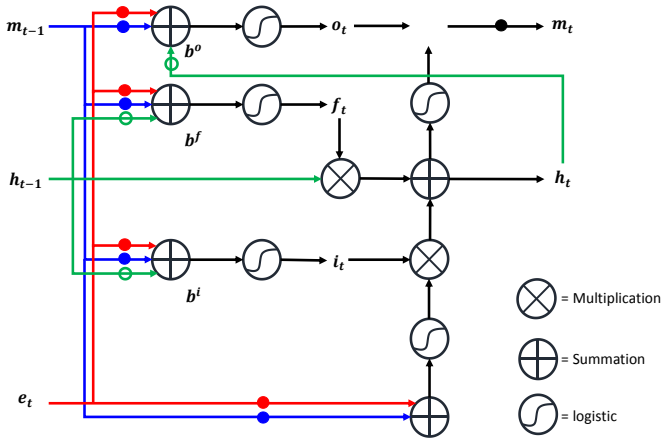


Figure 5.1: LSTM memory unit

$$m_t = W^{mm}(o_t \odot \sigma(h_t)) \tag{5.3}$$

$$o_t = logistic(\, W^{oe}e_t + W^{om}m_{t-1} + W^{oh}h_t + b^o) \tag{5.4}$$

$$h_t = f_i \odot h_{t-1} + i_t \odot \sigma(W^{he}e_t + W^{hm}m_{t-1} + b^h) \tag{5.5}$$

$$i_t = logistic(W^{ie}e_t + W^{im}m_{t-1} + W^{ih}h_{t-1} + b^i) \tag{5.6}$$

$$f_t = logistic(W^{fe}e_t + W^{fm}m_{t-1} + W^{fh}h_{t-1} + b^f) \tag{5.7}$$

$$e_t = W^{ex}x_t \tag{5.8}$$

where $\odot$ denotes element-wise multiplication and σ is sigmoid activation function. The W terms i.e. ($W^{fe}, W^{fm}, W^{fh}, W^{ie}, W^{im}, W^{ih}, W^{he}, W^{hm}, W^{oe}, W^{om}, W^{oh}, W^{mm}, W^{ym}$) are the models weights, and the b terms i.e. (b^f, b^i, b^h, b^o) are the bias terms. The weights and bias form the parameters ($[W, b]$ of the model, which are obtained during training. Training and optimization of the model is performed using Adam optimization [232] with a fifth of the training samples randomly chosen for each epoch.

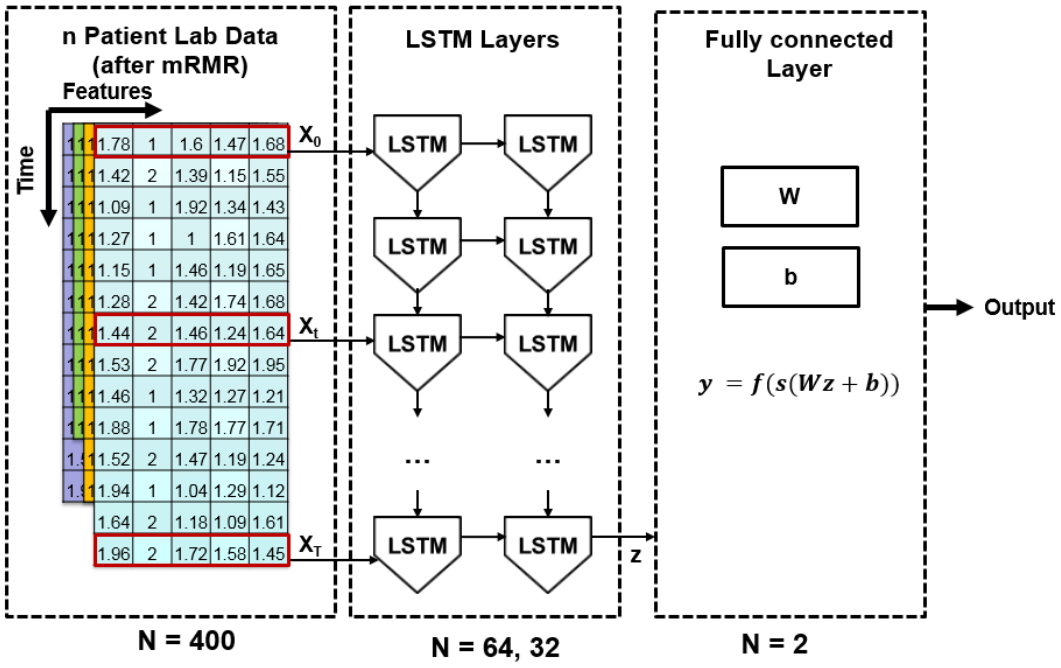


Figure 5.2: LSTM network. The patient features after feature selection is passed through LSTM layer and the intermediate features so generated are then passed through a fully connected layer for temporal analysis.

Each LSTM layer consists of $t = 1 - T$ LSTM units joined and evaluated iteratively (Figure 5.2). Our LSTM network consists of multiple LSTM layers, with the number of hidden layer in every successive layer half the previous layers. The decision made at the last time point is the final output of the LSTM network. For temporal prediction, the results of the last time point available for each patient is passed via a linear layer followed by a softmax layer. Hyper-parameters such the number of layers and size of each layer was estimated using the performance on a validation dataset. We tried 1 ,2 LSTM layers with the size of the first layer being 128, 64, 32, and 16.

5.2.5 Integration of Temporal and Non-Temporal Models

The integration of temporal and non-temporal models are performed through the integration of intermediate features generated using the individual models (Figure 5.3). Intermediate features are generated from the individual models when the input is passed

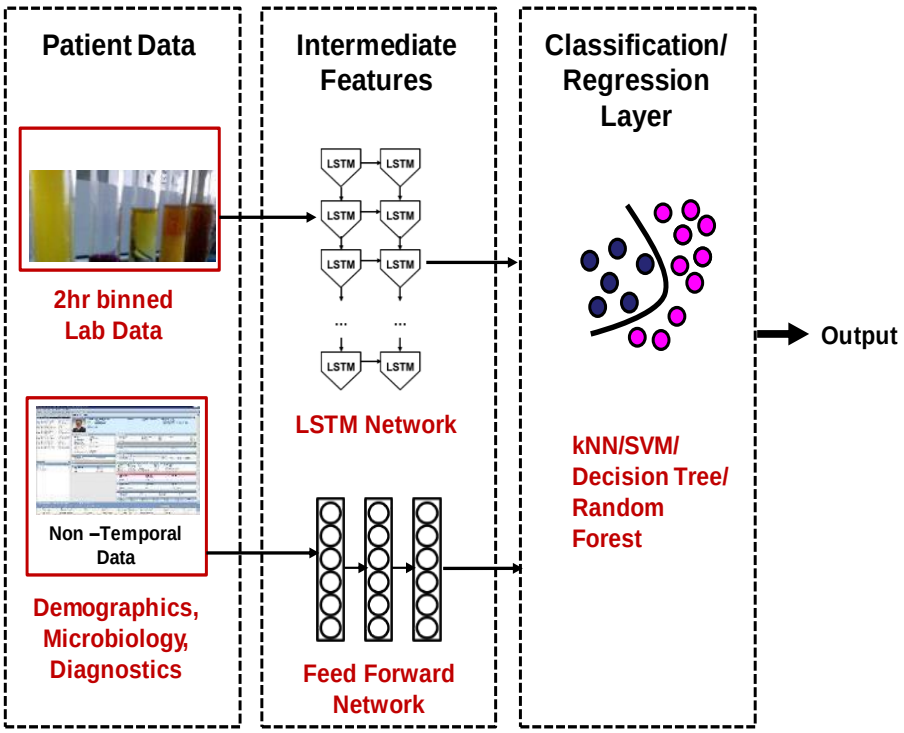


Figure 5.3: Integration of temporal and non-temporal data.

through all layers expect the final classification (linear + softmax layer). These features are concatenated using a linear layer followed by a classifier. The classifiers tested for the classification layer include k nearest neighbors, random forests, decision trees, and support vector machines.

5.2.6 Model Implementation

We developed the deep-models on torch lua environment. The libraries used included nn, nnx, optim, autograd and rnn. We ran the models on the pace clusters service by Georgia Tech (https://pace.gatech.edu/overview). Pace clusters can be accessed by writing a short proposal detailing the project and its usage by PIs. Each new student account also requires a description of the student's project, GT ID and a mandatory tutorial. Due to a lower number of GPUs as compared to the CPU, we trained our models on CPUs. We used 10 CPUs for training the deep models with each epoch taking about 2.5 hours. For the baselines, and classification layers, we used matlab2016b also in pace clusters.

5.2.7 Model Evaluation

From the total samples, 10% of the data was kept as external test set. Remaining data was used for training and validation using 3 fold cross validation. Cross validation was used for hyper-parameter evaluation. The temporal classification models were evaluated using GRUs (chosen since they are temporal deep models), and standard machine learning techniques such as k nearest neighbors, random forests, decision trees, and support vector machines. Standard machine learning techniques were used on aggregated data (means of the evaluation period). The non-temporal models were evaluated using k nearest neighbors, random forests, decision trees, and support vector machines. The integrated models were also evaluated against feature level integration techniques. The LOS

prediction using a combination of deep models with regression layer was evaluated using linear, SVM, decision trees, and random forests regression as baselines The evaluation metrics for classification include accuracy, Matthews correlation coefficient, precision, recall and F1 scores [233, 234]., and for regression include RMSE and MAE.

5.2.8 Model Interpretation using Perturbation Analysis

Interpretability of deep-learning models constitutes a major challenge. Rigorous research on the understanding of the representations coded by deep models has led to the conclusion that the behavior of deep models are very complex and a direct interpretation of the units and layers in the network can give rise to erroneous conclusions [235]. As an alternative, predictions/ conclusions from deep layers are passed as inputs to other machine learning models for improving interpretability [235]. Since in EHR data analysis, interpretation of features is of particular importance, in this work, we interpret the models by masking each feature in the classification model and checking for the changes in MCC.

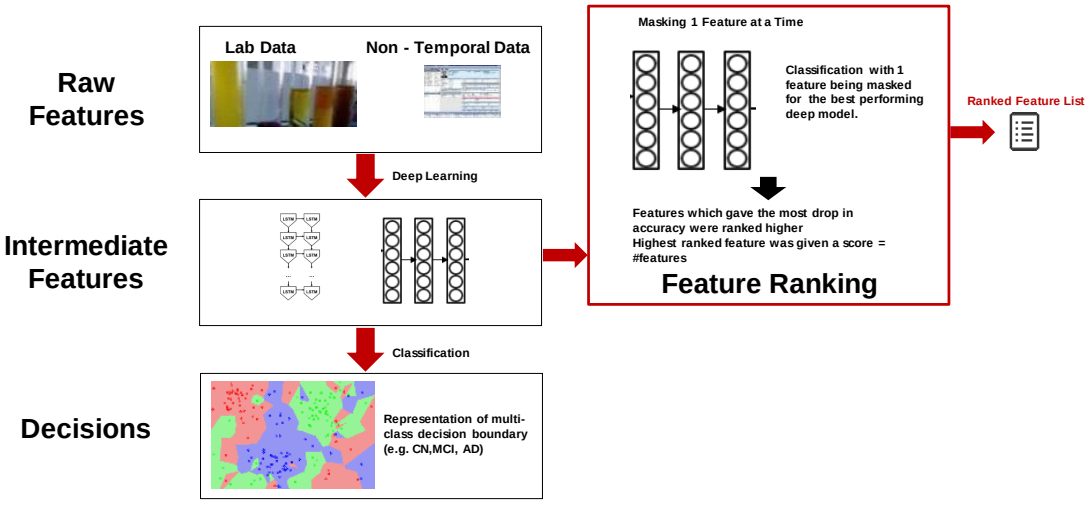


Figure 5.4: Model interpretation pipeline. The features for the deep models are masked one at a time and the effect on the classification is observed. The feature which gives the highest drop in accuracy is ranked the highest. Once we ranked the features, we checked if the intermediate picked associations different from raw data using cluster analysis.

The features which gave the highest drop in MCC was ranked higher (Figure 5.4.). We report the top features which gave the highest drop in MCC on the test sets.

5.3. Results

As mentioned above, we demonstrated our results using CHOA dataset with a total of 5,739 records. Patients had an ICU length of stay > 7 days in 2,174 records. We utilized up to 4 days of lab data (temporal) and demographics, microbiology, pathology, diagnostic codes and procedure details to classify which patients were likely to have LOS greater than seven days. We also predicted the length for ICU stay for patients with long LOS using regression models

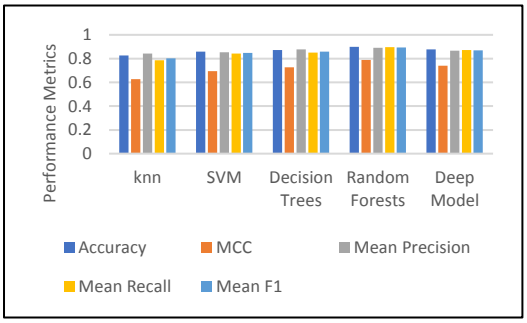
5.3.1 Classification of Patients at Risk of Length of Stay >7

Non-Temporal Analysis

We used a feedforward neural network with 1 input layer and 2 hidden layers, where the hidden layers had 200 and 100 nodes each. The training of the networks was done using Adam with a max epoch count of 10. The classification results in cross validation (Table. 5.2a.) indicate that the deep-learning models outperformed kNN, SVM and decision trees. Random forests gave the best performance. Despite this, we chose neural networks for integration due to the ease of integration and the ability to fine tune the integrated models. On the external dataset (Table 5.4), ventilator days, number of billable procedures, number of custom disease codes, number of ICD-9 codes, and the number of diseases coded final were the top features selected. Number of ventilator days [236, 237]; discharge destination [238, 239]; and ICU source [240-242] have been shown to be associated with longer ICU LOS.

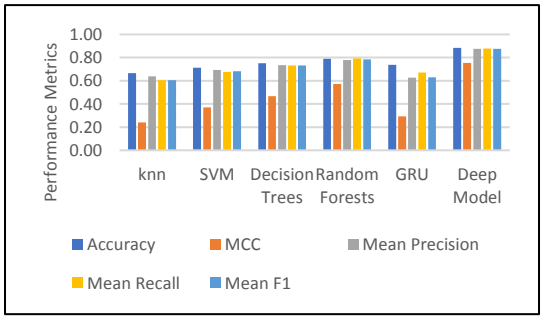
Table 5.2: Classification Cross Validation Result a) Non-Temporal Results b) Temporal Results c) Integrated Results. The kNN, SVM , RF and decision trees are baseline models. (kNN refers to k-nearest neighbors, SVM refers to support vector machines, RF refers to random forests, and IntF is the intermediate features which care combined with the different classification layers).

	kNN	SVM	Decision Trees	Random Forests	Deep-Method
Accuracy	0.83 ± 0.01	0.86 ± 0.01	0.87 ± 0.01	**0.9 ± 0**	0.88 ± 0.01
MCC	0.63 ± 0.02	0.69 ± 0.03	0.73 ± 0.02	**0.79 ± 0.01**	0.76 ± 0.03
Precision	0.84 ± 0.01	0.85 ± 0.01	0.88 ± 0.01	**0.89 ± 0**	0.87 ± 0.02
Recall	0.79 ± 0.02	0.84 ± 0.02	0.85 ± 0.01	**0.9 ± 0**	0.88 ± 0.01
meanF1	0.8 ± 0.01	0.85 ± 0.01	0.86 ± 0.01	**0.89 ± 0**	0.88 ± 0.01



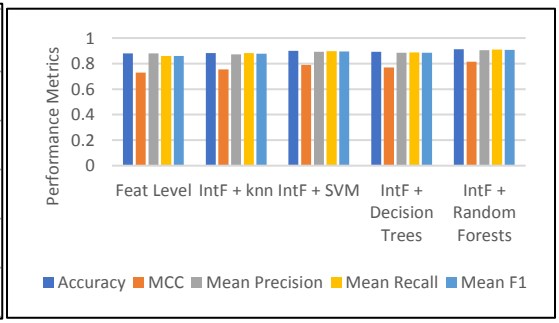
a: Results of non-temporal analysis. The results indicate that random forests outperformed deep models and all other baselines.

	kNN	SVM	Decision Trees	Random Forests	GRU	Deep-Method
Accuracy	0.67 ± 0.01	0.71 ± 0.02	0.75 ± 0.01	0.79 ± 0.01	0.74 ± 0.03	**0.88 ± 0.01**
MCC	0.24 ± 0.02	0.37 ± 0.05	0.47 ± 0.02	0.57 ± 0.01	0.29 ± 0.05	**0.75 ± 0.01**
Precision	0.69 ± 0.01	0.75 ± 0.02	0.8 ± 0.01	**0.87 ± 0**	0.63 ± 0.01	0.84 ± 0.01
Recall	0.59 ± 0.02	0.64 ± 0.03	0.67 ± 0.01	0.69 ± 0.01	0.67 ± 0.04	**0.91 ± 0.01**
meanF1	0.85 ± 0.01	0.82 ± 0.01	0.81 ± 0.02	0.78 ± 0.01	0.63 ± 0.01	**0.86 ± 0.02**



b: Results of temporal analysis. The results indicate that LSTMs outperformed all baselines (kNN, SVM, decision trees, random forests and GRU)

	Feat Level	IntF + kNN	IntF + SVM	IntF + Decision Trees	IntF + Random Forest
Accuracy	0.88 ± 0	0.88 ± 0.01	0.9 ± 0.01	0.89 ± 0	**0.91 ± 0.01**
MCC	0.73 ± 0.01	0.75 ± 0.01	0.79 ± 0.02	0.77 ± 0.01	**0.82 ± 0.01**
Precision	0.88 ± 0	0.87 ± 0.01	0.89 ± 0.01	0.88 ± 0	**0.9 ± 0.01**
Recall	0.86 ± 0	0.88 ± 0.01	0.9 ± 0.01	0.89 ± 0	**0.91 ± 0.01**
meanF1	0.86 ± 0	0.88 ± 0.01	0.89 ± 0.01	0.89 ± 0	**0.91 ± 0.01**



c: Results of integrated analysis. The results indicate that integrated models outperformed individual models and feature level integrations.

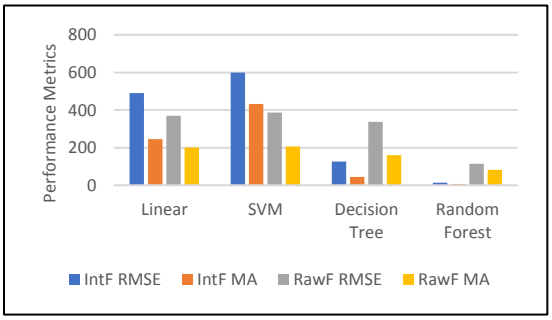
Temporal Analysis

After validation, we used a 2 layer LSTM network with 64 outputs nodes in the first hidden layer and 32 output nodes in the second hidden layer of the network. As

mentioned above the input to the network were the 400 selected features. The top features which were input into the model included the number of times the cyclosporin, calcium, heparin assay, amikacin trough, tacrolimus, state metabolic screen, tobramycin random,

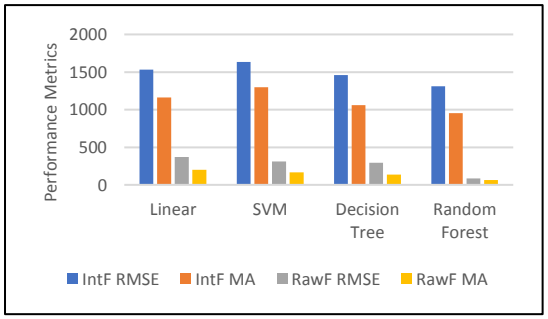
Table 5.3: Regression Cross Validation Result a) Non-Temporal Results b) Temporal Results c) Integrated Results. The linear, SVM, decision trees, and random forests regression models which are run on the intermediate features generated using deep models and on raw features. (IntF is the intermediate features and RawF are raw features).

	Input Type	Linear	SVM	Decision Tree	Random Forest
RMSE	Intermediate Features	490.72 ± 6.92	598.72 ± 5.61	125.46 ± 23.75	**13.83 ± 0.4**
MAE		244.9 ± 5.26	433.19 ± 2.08	44.81 ± 3.77	**7.56 ± 0.09**
RMSE	Raw Features	369.93 ± 3.33	387.27 ± 28.62	336.87 ± 59.01	**114 ± 1.58**
MAE		201.12 ± 2.19	205.82 ± 10.56	160.64 ± 7.31	**82.02 ± 1.35**



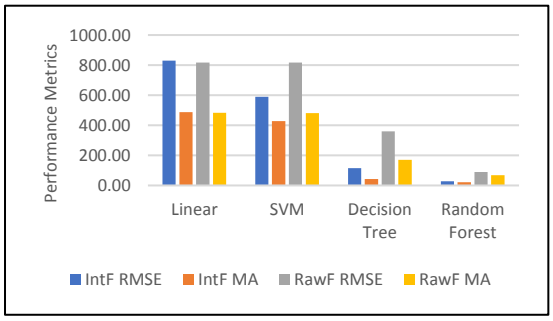
a: Results of non-temporal analysis. The results indicate that random forest regression with intermediate features outperformed deep models and all other baselines.

	Input Type	Linear	SVM	Decision Tree	Random Forest
RMSE	Intermediate Features	1533.91 ± 4.75	1635.67 ± 2.6	1459.8 ± 3.96	**1311.51 ± 10.35**
MAE		1160.38 ± 4.1	1299.46 ± 1.23	1061.12 ± 6.27	**951.62 ± 8.53**
RMSE	Raw Features	824.65 ± 58.55	817.49 ± 60.17	782.53 ± 59.09	**673.45 ± 62.55**
MAE		486.79 ± 7.13	479.38 ± 7.2	446.49 ± 7.47	**362.04 ± 13.93**



b: Results of temporal analysis. The results indicate that random forest based regression models with raw features outperformed all other models.

	Input Type	Linear	SVM	Decision Tree	Random Forest
RMSE	Intermediate Features	829.03 ± 57.49	589.16 ± 3.7	115.68 ± 10.72	**28.37 ± 0.33**
MAE		486.32 ± 7.04	427.55 ± 1.71	42.62 ± 2.13	**21.55 ± 0.49**
RMSE	Raw Features	817.41 ± 63.4	816.26 ± 60.54	358.92 ± 51.8	**88.55 ± 3.88**
MAE		482.21 ± 7.81	481.77 ± 7.32	169.31 ± 2.54	**67.9 ± 2.31**



c: Results of integrated analysis. The results indicate that random forests based regression models with intermediate features outperformed all other models.

soluble IL-2R, IGG, and lidocaine were performed. The training of the networks was done using Adam with a max epoch count of 10. For each of the epochs, 1000 samples of the training data was selected at random. The cross-validation results (Table. 5.2b.) indicate that the deep-learning EHR models outperformed all the standard models and GRUs.

On the external test set (Table 5.4.), capillary POC pH tests, the number of times the specimen was taken, methemoglobin tests, mean platelet volume tests and drug screen serum tests were the top features. Blood pH changes [243] [244]; methemoglobinemia [245] [246]; and hemoglobin [247, 248] have been shown to be associated with longer ICU LOS.

Integrated Models

The integrated models used the intermediate features from non-temporal (100 features) and temporal (32 features), which were added to a classification layer. We tried kNN, SVM, decision trees and random forests for the classification layers. Results (Table 5.2c.) indicate that the integrated models outperformed either of the individual modalities for ICU length of stay prediction. The deep models with the classification layer random forests outperformed individual models and feature level combinations.

For imbalanced datasets such the EHR data, MCC is a better measure than of performance than accuracy [233]. When we compare the MCC of the best performing combined model with the best performing temporal model, we see a 7% change on the external test set, with the integrated models performing better. Similarly, we observed a 3% improvement over non-temporal models.

On the external test set (Table 5.4.), number of times flu A,B RSV PCR was performed, the number of times GRAM STAIN test (check the presence of gram negative bacteria) was performed, the number of times the specimen source was abdominal fluid,

and the number of times the specimen source was Abscess. The presence of gram negative bacteria [249] [250]; abdominal fluid abnormalities [251] [252]; and abscess [240, 253]

Table 5.4: Top 10 features from classification and regression analysis. The lab data consisted of information on the tests and procedures conducted (labeled as component name along with the procedure name or the just the test name), the source of specimens (e.g. blood serum, urine and labeled as source), and the number of abnormalities in tests and procedures performed (labeled as Result status)

SN	Integrated Model	Imp 90% features = 6560	Temp	Imp 90% features = 360	NT	Imp 90% features = 699
1	PROC NAME FLU A,B RSV PCR	0.838984214	COMPONENT NAME CAPILLARY POC PH	0.828559586	PICU YN	0.725063147
2	PROC NAME GRAM STAIN	0.838984214	# of TIMES SPECIMEN TAKEN	0.816858229	NICU YN	0.708318298
3	SPECIMEN SOURCE Abdomen	0.838984214	COMPONENT NAME METHEMOGLOBIN	0.813789998	CICU YN	0.655478928
4	SPECIMEN SOURCE Abdominal fluid	0.838984214	COMPONENT NAME MEAN PLT VOLUME	0.811892576	FINANCIAL CLASS GROUP	0.651155848
5	SPECIMEN SOURCE Abscess	0.838984214	SOURCE SYSTEM PROC CODE SDRUG	0.810563857	FIN CLASS	0.645038011
6	SPECIMEN SOURCE Anal	0.838984214	SOURCE SYSTEM PROC CODE CSFPRO	0.810563857	PRIMARY PAYOR	0.642208212
7	SPECIMEN SOURCE Arm	0.838984214	COMPONENT NAME HGB A (ELECTRO HGB)	0.810563857	VENTILATOR DAYS	0.614632196
8	SPECIMEN SOURCE Arm left	0.838984214	SOURCE SYSTEM PROC CODE DICSCR	0.810379421	ADMISSION SOURCE	0.612274664
9	SPECIMEN SOURCE Arm right	0.838984214	SOURCE SYSTEM PROC CODE OIA	0.809422574	TRANSFER SOURCE	0.610305945
10	SPECIMEN SOURCE Arterial	0.838984214	SOURCE SYSTEM PROC CODE PT	0.809422574	DISCHARGE DESTINATION	0.592021843

a) Results of classification analysis.

SN	Integrated Model	Imp 10% features = 1579	Temp	Imp 10% features = 41	NT	Imp 10% features = 1186
1	PICU YN	85.73504735	COMPONENT NAME POC HEMATOCRIT	282.1163408	PICU YN	119.9731743
2	COMPONENT NAME CAPILLARY POC PH	122.4905978	COMPONENT NAME MEAN PLT VOLUME	282.9276644	NICU YN	249.2230445
3	SOURCE SYSTEM PROC CODE GENTT	122.6107445	COMPONENT NAME MCHC	283.2248576	CICU YN	276.32218
4	COMPONENT NAME POC GLUCOSE	124.4368015	COMPONENT NAME POC SODIUM	283.7647028	VENTILATOR DAYS	388.9507035
5	SOURCE SYSTEM PROC CODE T4FREE	125.3644974	COMPONENT NAME CAPILLARY POC PH	283.9915591	TRANSFER SOURCE	401.6700675
6	SOURCE SYSTEM PROC CODE AMY	126.6121282	COMPONENT NAME MCV	283.9995022	SEX	410.9654217
7	COMPONENT NAME POC ACT COAG TIME	126.8693052	SOURCE SYSTEM PROC CODE UAMIC	284.9210802	ADMISSION SOURCE	411.4862638
8	SOURCE SYSTEM PROC CODE POCAI	126.9729629	RESULT STATUS Edited	284.9379048	DISCHARGE DESTINATION	411.5138283
9	SOURCE SYSTEM PROC CODE AMON	127.1179185	COMPONENT NAME TOTAL PROTEIN	285.2830284	RACE	538.9405466
10	COMPONENT NAME URINE RBC	127.3371111	COMPONENT NAME CSF PROTEIN	285.312982	FINANCIAL CLASS GROUP	541.1467835

b) Results of regression analysis.

has been associated with longer ICU LOS.

5.3.2 Regression Analysis for Predicting Length of ICU Stay

For patients at risk for LOS>7, we performed a regression test for predicting the actual length of stay. We tried the intermediate features generated using deep-models in the previous step against the raw features for predicting LOS using linear, SVM, decision trees, and random forests based regression techniques

Non-Temporal Analysis

The regression results in cross validation (Table. 5.3a.) indicate that random forest regressions using intermediate features outperformed all other base lines and those using intermediate features. The external validation results (Table 5.4) showed that the non-

Table 5.5: External Test Result. For classification deep models outperformed for temporal and integrated analysis. For regression, deep models outperformed for non-temporal and integrated analysis.

	Models	**External Test Performance**
Temporal Analysis Classification	LSTM: #features 400, Layer sizes (64, 32) Classification Layer: Random Forest Trees = 498	Accuracy: 0.88 MCC: 0.76 Precision: 0.88 Recall: 0.88 F1 Scores: 0.88
Non-Temporal Analysis Classification	Random Forest Trees = 100;	Accuracy: 0.89 MCC: 0.77 Precision: 0.88 Recall: 0.88 F1 Scores: 0.88
Integrated Analysis Classification	Deep-Model: Non-Temporal Layer sizes (200, 100) LSTM: #features 400, Layer sizes (64, 32) ; classification layer = Random Forest Trees = 496;	Accuracy: 0.91 MCC: 0.81 Precision: 0.90 Recall: 0.91 F1 Scores: 0.91
Temporal Analysis Regression	Random Forest Trees = 234	RMSE: 673.70 MAE: 411.34
Non-Temporal Analysis Regression	Deep-Model: Layer sizes (200, 100); regression layer = Random Forest Trees = 492	RMSE: 221.54 MAE: 82.43
Integrated Analysis Regression	Deep-Model: Non-Temporal Layer sizes (200, 100) LSTM: #features 400, Layer sizes (64, 32) ; regression layer = Random Forest Trees = 425	RMSE: 240.36 MAE: 122.46

temporal analysis results were the best. On the external dataset (Table 5.4), ICU visit (Y/N) (for PICU, NICU, and CICU); number of ventilator days [236, 237]; discharge destination [238, 239]; gender [254, 255], glucose abnormalities [256, 257] and ICU source [240-242] are top features which have been shown to be associated with longer ICU LOS.

Temporal Analysis

The regression results in cross validation (Table. 5.3b.) indicate that random forest regressions using raw features outperformed all other base lines and those using intermediate features. The external validation results (Table 5.4) also showed that the temporal analysis results using only the details of the test performed and the number of abnormalities were not as informative as the non-temporal (with demographics, medication, procedure and ICD-9 codes) and integrated models for the prediction of ICU LOS. On the external test set (Table 5.4.), hematocrit tests [258, 259]; sodium tests [260, 261]; and blood pH changes [243] [244] are top features which have been shown to be associated with longer ICU LOS.

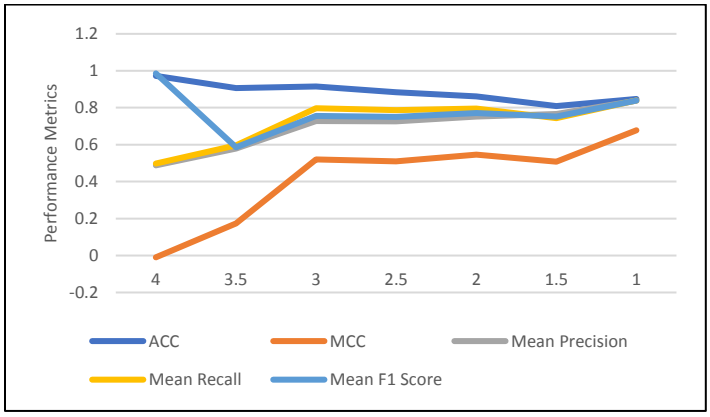
Integrated Analysis

The regression results in cross validation (Table. 5.3c.) indicate that random forests regressions using intermediate features outperformed all other base lines and those using intermediate features. The cross validation and the external validation results (Table 5.4) showed that the integrated models with deep learning generated intermediate feature gave an improved performance for LOS over temporal models. This shows the synergistic effects of the integrated models. On the external test set (Table 5.4.), act coagulation time[262, 263]; urine RBC tests[258, 259]; blood pH changes [243] [244]; and

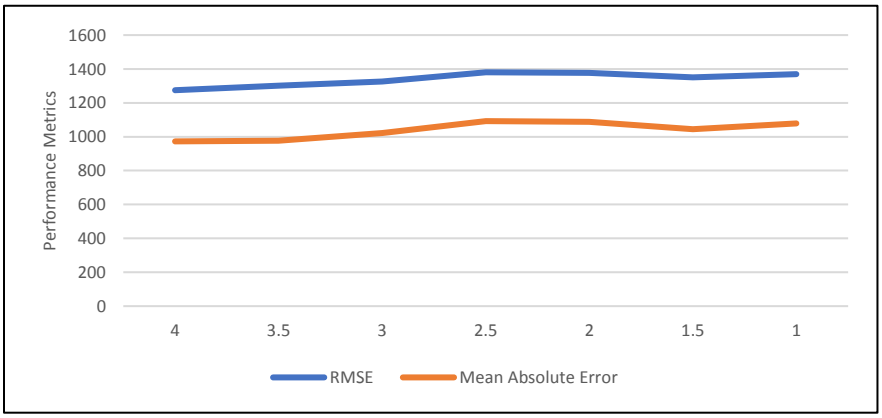
cerebrospinal fluid protein tests [264, 265] are top features which have been shown to be associated with longer ICU LOS.

5.3.3 Analysis of Temporal Data for Effects of Multiple Time Windows (Long - Term Memory Component)

In order to test the effects of using longer term data as opposed to short term data, we used a window based approach for the LSTM models. We started backwards with using only the fourth day data. Then we progressively increased the window size by adding half a day word of data at each step until we reached the end. We performed this analysis for both the classification models. For classification, our results (Figure. 5.5a.) indicate that the prediction metrics MCC, mean precision, mean recall and mean F1 score steadily increased with increased wind size and was the highest when all the data included the first time step information was added. This analysis shows the LSTM models trained for LOS risk prediction requires long term dependencies for improved decision making as opposed to short term dependencies alone. Conversely, for the regression analysis results (Figure. 5.5b.), the increasing the window decreased the performance. The RMSE and MAE values increased when longer term information was used. This shows that the prediction of LOS placed higher dependencies on the short term features as opposed to long term relationships in data. This is also supported by regression analysis using early versions of LSTMs [266]



a. Classification Results Windowing Analysis: We used upto 4 days of information to predict length of stay. We added details of .5 days at a time. (only 3-4 days, 2.5-4 days, 2-4 days, 1.5-4 days, 1-4 days, .5-4 days and 0-4 days). Adding more long-term details improved performance measured using MCC, mean precision, mean recall, mean F1 scores and accuracy



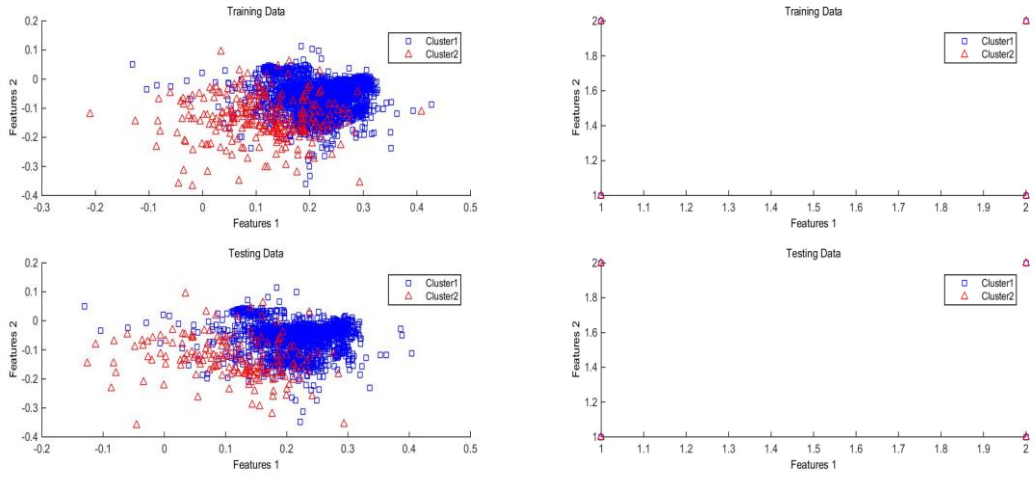
b. Regression Results Windowing Analysis: We used upto 4 days of information to predict length of stay. We added details of .5 days at a time. (only 3-4 days, 2.5-4 days, 2-4 days, 1.5-4 days, 1-4 days, .5-4 days and 0-4 days). Adding more long-term details did not change performance much but dropped performance slightly. Performance was measure using RMSE and MAE

Figure 5.5: Windowing Analysis. a) Classification analysis b) Regression Analysis (MCC refers to Matthews correlation coefficient, RMSE refers to root mean square error and MAE refers to mean absolute error).

where window based approaches using smaller windows gave a better performance for regression problems as opposed to LSTM networks. In addition, small data sizes may affect the performance of LSTMs for regression as well.

5.3.4 Analysis of Intermediate Features from for Data Associations not Seen in Raw Data.

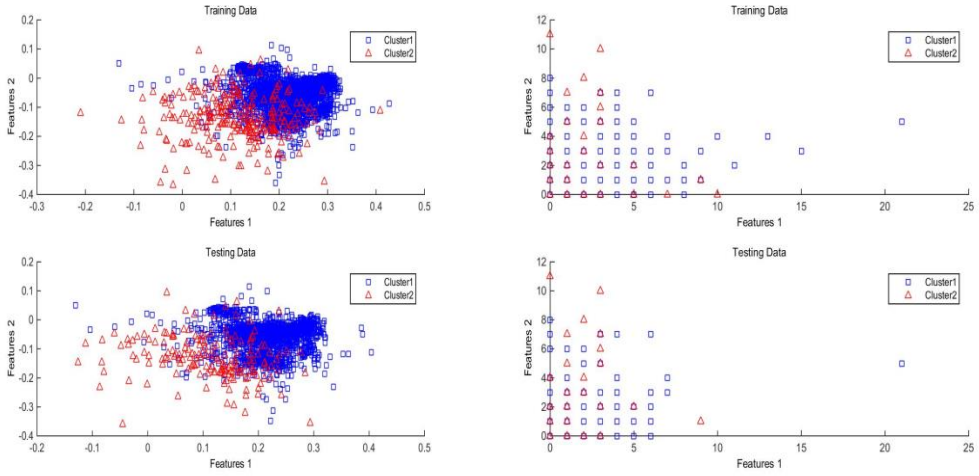
In addition to the tests for the effects of short term and long term dependencies on the model performance, we also evaluated the intermediate features generated from the deep models using a cluster analysis. This was performed to check for data relationships picked by the deep models which are not easily discernible in the raw features. For this analysis, we clustered the intermediate features (training set) from temporal and non-temporal models using kmeans. We used the training data to fix the number of clusters and



a b

Figure 5.6: Cluster analysis results: Temporal data. a) gives the intermediate features plotted against the top ranked features which discriminated the two clusters using relieFF. b) gives the original data plotted against the top ranked features which discriminated the two clusters using relieFF

cluster centers. Then we map the test data to the same cluster centers. Then we check if the separation of intermediate features into clusters or the lack is consistent across training and test samples. The cluster number was evaluated using the mode of cluster number generated using Calinski Harabasz [267], Davies-Bouldin [268], silhouette [269] and gap scores [270]. For both the non-temporal and temporal data, we clustered the intermediate



a b

Figure 5.7: Cluster analysis results: Non-Temporal data. a) gives the intermediate features plotted against the top ranked features which discriminated the two clusters using relieFF. b) gives the original data plotted against the top ranked features which discriminated the two clusters using relieFF

features generated. We test a cluster number of 2-20 to pick the optimum cluster number, which was found to be two. We plotted the intermediate features and the original data against the top features (relieFF) which discriminated the two clusters (Figure 5.6a,b, 5.7a,b). We found that the intermediate features generated better separation as compared to the original data. This indicates that the intermediate features found relationships which are not easily apparent in original data.

In summary, ours results indicate the following

- For classification of patients into those with a risk for long length of ICU stay, LSTM based models gave best performance for temporal data alone, random forest based models gave an improved performance for non-temporal data alone. In the presence of both data types, the integrated models using LSTM and feed forward neural networks gave the best predictive performance.

- For the prediction of numerical LOS, deep models were beneficial when non-temporal data available and when both data types were available. With only temporal data type, random forest based techniques on raw data gave the best performance.
- To determine risk of long LOS, the long term temporal dependencies give improved predictive performance. To the prediction of numerical LOS, short term dependencies were found to be more meaningful.

5.4. Conclusion and Key Innovations

Most common models for ICU length of stay (LOS) are based on the use of static or temporal data In this study, developed deep-learning strategies for the integration of static and temporal ICU data. We utilized an LSTM based approach for temporal data and neural networks for non-temporal data. The combination was performed using a classification layer/ regression layer. We evaluated our models on pediatric data from CHOA by comparing our models using gated recurrent units, random forests, support vector machines, k nearest neighbors, and decision trees. Using accuracy, Matthews correlation coefficients, precision, recall, and mean F1 scores as evaluation metrics, our integrated models outperformed all baselines for predicting long ICU stay. The integrated models using deep models also outperformed the shallow learned and the individual modalities for the prediction of the actual length of ICU stay.

However, despite good prediction, our model suffers from the challenges in interpretability. In this analysis, we address this partially using a perturbation-based approach to interpret the intermediate features generated using deep learning. In the future,

we will extend this work to include more interpretability by following a sequential feature selection based approach and adding features sequentially and a back propagation based method. In addition, in our current work, we used the intermediate feature generated for classification to perform regression. Though this shows the generalizability of the features extracted, we may be able to show improvements in performance by using a dedicated deep regression model in the future. We will also investigate the use of more window based regression techniques such as ARMA and ARIMA [266] in our analysis. We will also extend our analysis to include more end-points such as ICU readmission, and ICU mortality.

We will also investigate the use of more sophisticated feature section techniques such as mutual information with missing data [228], margin based feature selection [229], time series feature[271] selection which is robust to missing time series data.

To summarize, the key innovations of this chapter include:

- We developed a framework for combining data from multiple temporal resolutions using deep models.
- We demonstrate ICU length of stay prediction on pediatric ICU patient data
- We developed perturbation based analysis for data interpretation.
- We showed the use of long-term dependencies for patient classification.

CHAPTER VI

DEEP LEARNING MODELS FOR INTEGRATING ELECTRONIC HEALTH RECORD DATA WITH GENETIC DATA FOR ALZHEIMER'S DISEASE PREDICTION

6.1. Introduction

Alzheimer's disease (AD) is the most common neurodegenerative disorder [272] and forms the 6th leading cause of death in the United States. There are more than 5.3 million people living with Alzheimer's in the United States alone [273, 274]. In addition, the mortality caused by AD has steadily increased over past 30 years, contributing to over 83,494 deaths in 2010 [275]. The healthcare cost for AD is also steadily increasing. It was estimated to be 200 billion dollars in 2012 and is expected to be 1.1 trillion dollars by 2050 [275]. Despite extensive research and advances in clinical practice, there is still a lack of complete knowledge regarding the biomarkers which are indicative of AD and its stages. Less than 50% of the people with AD are being diagnosed accurately for their pathology and disease progression on the basis of their clinical symptoms [273]. The presence of amyloid plaques and neurofibrillary tangles in histopathology is the most conclusive evidence for Alzheimer's diagnosis. However, the early onset of AD is not correlated with the presence of plaque but with synaptic and neuronal loss [276].

Research on data from Alzheimer's disease initiative [277] and data mining strategies [278-282] for AD are being undertaken to improve our understanding of the underlying disease processes. AD biomarkers including clinical symptoms (such as dementia, memory loss), neurological tests [283] and scores such as MMSE scores[284] are being augmented with several imaging, genetic and protein related biomarkers.

Cerebrospinal fluid-based biomarkers such as Aβ1-42 levels (indicative of amyloid deposition in the brain) [285], total tau protein and hyperphosphorylated tau protein [286], YKL-40 chromogranin A, and carnosinase I [287] have been shown to be well-accepted markers indicative of early and advanced AD. Similarly imaging markers such as white matter hyperintensities [10], volume reductions of the medial temporal lobes (indicative of neuronal loss and accumulation of neurofibrillary tangles), pathological involvement of limbic and cortical regions [288], and changes in hippocampus volume and structure [289] have been identified using positron emission tomography (PET) and magnetic resonance imaging (MRI) studies. Research is being undertaken on blood-based biomarkers such as blood proteins, circulating miRNAs, small endogenous RNAs [290] to find relatively non-invasive biomarkers. However, most of these studies identify biomarkers using a single modality or data type. This limits the ability of the data mining algorithms to find more generalized biomarkers and subsequently the mechanisms indicative of AD. Use of such limited datasets (hence bio-markers) also restricts the performance of these algorithms for the early identification of AD disease and its stages. In addition, most of these studies perform binary classification into either AD/MCI vs controls or AD vs controls.

Current multi-modal analysis for AD mostly combines various imaging modalities [291-295] such as structural MRI (T1 weighted, T2 weighted, DTI, DWI), fMRI and PET [296, 297]. Another established multi-modal analysis for AD is imaging genetics [298]. However, these studies mainly use traditional machine learning techniques such as t-tests [292], support vector machines (SVMs) [291], and principal component analysis (PCA)[299], which may not take full advantage of integration of different data. In addition, these techniques fail when the data for particular modalities are absent. The recent

advances in deep learning [300] and their applications to Alzheimer's imaging data [301-303] have shown promising results in improving upon the prediction power of the AD models. Deep learning based image fusion studies using PET and MRI data also report improved predictive performance with auto-encoders [304, 305], and deep-belief networks [306]. Deep-learning studies for EHR [120] and SNP [307] data have also shown improvement over traditional machine learning. In addition, the use of deep-learning techniques also facilitates the training and prediction in the presence of partial data [308].

In this study, we further the multimodal AD data fusion to advance AD stage prediction by using DL to combine imaging, electronic health record and genomic SNP data for the classification of patients into control, MCI, and AD group. We use stacked de-noising auto-encoders for EHR and SNP data respectively, and novel 3D convolutional neural networks to train MRI imaging data. After the networks are separately trained for each data modality, we combine them using different classification layers including decision trees, random forests, support vectors machines (SVM) and k-nearest neighbors (kNN). We demonstrate the performance of our integration models using the ADNI [309] dataset that contains SNP (808 patients), imaging (MRI) data (503 patients), and clinical and neurological test data (2,004 patients).

Despite superior performance in clinical decision support using multiple data types, a major drawback for widespread adoption of DL models for clinical decision making is the lack of well-defined methods for interpreting the deep models. We address this challenge by developing novel perturbations and clustering-based approach for finding the top features contributing to the decision.

In this study, we report the major contributions for the AD stage prediction as follows :

- Novel DL architectures outperform shallow learning models;
- Multi-modality data analysis with DL outperforms single-modality DL models; and
- Novel interpretable DL methods are capable of extracting top performing features.

6.2. Methods

6.2.1 Data Description

This study uses Alzheimer's Disease Neuroimaging Initiative* (ADNI) database (adni.loni.usc.edu) [309] data for the analysis. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessments can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). ADNI data repository contains imaging, clinical and genetic data for over 2,220 patients spanning over 4 studies (ADNI1, ADNI2, ADNI GO, and ADNI3). In our study, we focus on ADNI1, 2 and GO because ADNI 3 is an ongoing study which is due to end in 2022. The data is currently being released in phases with limited availability for imaging (unprocessed) and no genetic data yet. The imaging data (ADNI1, 2 and GO) consists of MRI and PET images, of which we use cross-sectional MRI data corresponding to the baseline screenings from ADNI1 (503 patients). The images have been standardized by the publisher of the data to eliminate the non-linearities caused by the scanners from different vendors. For clinical or EHR data, we use 2,004 patient (ADNI1, ADNI2, and ADNI GO) data from the clinical tests (e.g. memory tests, balance

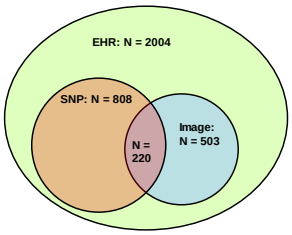
Table 6.1: 1a. Description of ADNI data. Clinical data consists of demographics, neurological exams and assessments, medications, imaging volumes and biomarkers. 1b Number of patients by modality and disease stage. (CN: controls; MCI: Mild Cognitive disorder and AD: Alzheimer's Disease). 1c Venn diagram showing the degree of overlap between the three modalities. 220 patients had all the three data modalities, 588 patients had SNP and EHR, 283 patients had imaging and EHR, the remaining patients had only EHR data.

	Example Data Types/ Features
Clinical Data	Demographics, neurological exams, cognitive assessments, bio-markers (e.g. alanine, choline), medication (e.g. levodopa), imaging summary scores (e.g. brain are volumes)
Imaging	Cross-sectional MRI data
Genetic	Whole genome sequencing (WGS) data

a:

	CN	**MCI**	**AD**
Clinical Data	598	699	707
Imaging	132	104	266
Genetic	245	338	226

b:



c:

tests, cognitive tests), medication data (e.g. usage of levodopa), imaging score summaries (e.g. levels of FDG from PET, brain volumes from MRI), patient demographics (e.g. age, gender), and biochemical tests. The genetic data we use consists of the whole genome sequencing (WGS) data from 808 ADNI participants. The WGS has been performed on 818 subjects (at the time of sequencing, 128 with AD, 415 with MCI, 267 controls and 10 of uncertain diagnosis) from the ADNI study by Illumina's non-CLIA laboratory at roughly 30-40x coverage in 2012 and 2013. The resulting variant call files (VCFs) have been generated by ADNI using Broad best practices (BWA and GATK-haplotype caller) in 2014. Hence for this study, we use a total 2,004 patients for whom clinical data was

Table 6.2: The mapping rules for labels with disease progression

After mapping	CN Label = 1			MCI Label = 2			AD Label = 3		
Original	Stable CN to CN Label = 1	Reversion MCI to CN Label = 7	Reversion AD to CN Label = 9	Stable MCI to MCI Label = 2	Conversion CN to MCI Label = 4	Reversion AD to MCI Label = 8	Stable AD to AD Label = 3	Conversion MCI to AD Label = 5	Conversion CN to AD Label = 6

available, 503 patients with imaging data (9108 voxels per patient distributed over 18 slices, with each slice having 22×23 voxels), and 808 patients with genetic data (Table 6.1.). For participants with multiple visits, we use the diagnosis from patient's last visit. As shown in Table 6.1c., 220 patients have all three data modalities, 588 patients have SNP and EHR, 283 patients have imaging and EHR, the remaining patients have only EHR data

Most participants in ADNI1/GO/2 studies have multiple visits during the 48 month study period. The diagnosis of some of the patient's changes along time. However, in this study, we only focus on predicting the risk of developing AD instead of the progression of AD. To remove the influence of disease progression, we match the labels based on following rules: We only use the diagnosis of each patient's last visit. This will give us the latest label of the patient.

For labels representing the progression, we map them to following three corresponding labels: (1) control (CN), (2) mild cognitive impairment (MCI), and (3) Alzheimer's disease (AD). The detailed label mapping rules are summarized in Table 6.2.

6.2.2 Data Pre-processing

As mentioned above, ADNI dataset consists of MRI imaging data, clinical data, and SNP data. For each data modality, we perform feature extraction and selection respectively.

MRI Imaging Data

We first preprocess the 3D images to filter noise, perform skull stripping, segment different types of brain tissue, normalize and co-register the images to MNI space (Fig 4a.) [310]. Following that, we extract 3D areas of 21 brain regions (associated with Alzheimer's disease) including the right amygdala, left and right angular, left and right cerebellum, left and right Hippocampus, left and right occipital regions, and left and right superior temporal regions (Supplementary material).

Clinical Features

From ADNI1, ADNI2, and ADNI GO, we extracted common fields, from which we extracted 1,680 features. The clinical features found in this dataset were either quantitative real numbers, binary and categorical. We normalized the quantitative data to the range 1-2 and converted the categorical data into binary using one hot encoding. We also converted all the binary data into values 1 or 2.

Genetic Data

Each subject has about ~3 million SNPs in the raw VCF file. To eliminate irrelevant and redundant SNPs, we apply multiple filtering and feature selection steps. We first eliminate SNPs with 1) low genotype quality, 2) low minor allele frequency, 3) high per site missing rate, and 4) significant Hardy-Weinberg equilibrium p value. After filtering, we apply a two-stage feature selection. In the first stage we only retain SNPs that located on known AD associated genes. In the second stage, we further reduce the number of SNPs using minimum redundancy maximum relevance (mRMR) [198]. After feature extraction and selection, we obtain 500 SNP features for further analysis. Following feature selection

and reduction, we proceed to perform classification of the data into AD, MCI, and controls.

6.2.3 Intermediate Feature Generation using Individual Modalities

After feature selection, we use deep-learning techniques for the generation of intermediate features. The intermediate features generated from the individual modalities are subsequently used for multi-modal analysis. The intermediate features from EHR and SNP data are generated using auto-encoders.

Intermediate Features for EHR and SNP Data using Auto-Encoders

Each patient data (EHR and SNP) is represented as a vector of length m (where m is the number of features), and is used as an input to the feature learning algorithm. This data is then passed through a two-layer stacked denoising auto-encoder network [311] (Figure 6.1) to obtain a high level representation of the patient data. Each auto-encoder layer takes an input x of dimension $n \times d$, where n is the number of training samples and d is input dimensionality ($d = m$ for first layer). The input for each layer is first passed through an encoder to convert the input into a higher order representation of the data (6.1).

$$y = f(Wx + b) \quad (6.1)$$

where f is an activation function such as sigmoidal or tanh, $[W, b]$ are parameters to be trained. The mapped values (y) are then passed through a decoder to obtain a representation of the input(x) (6.2).

$$\hat{x} = f(W^T y + b') \quad (6.2)$$

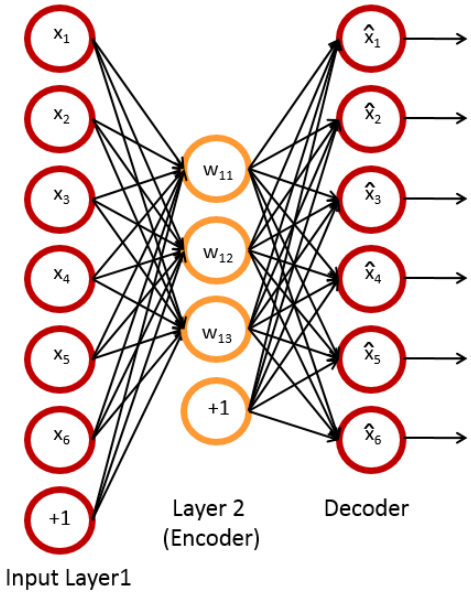


Figure 6.1: Auto encoder layers

where b' needed to be trained, and the weights W^T are tied with the encoder weights. The network is constructed by stacking the trained encoder layers. Denoising is implemented using dropouts where a portion of the input values are masked (set to zero) to allow better generalization of the models in the presence of small and noisy training data. Training is performed by back propagation by minimizing the average cross-entropy between the input and the reconstructed input data (6.3).

$$[W, b, b'] = \underset{[W,b,b']}{\arg min} - \sum_{k=1}^{a} [x_k log \hat{x}_k + (1 - x_k)\, log(1 - \hat{x}_k)] \quad (6.3)$$

where a is number of dimensions. Optimization is carried out using Adam optimization [232]with a batch size of three.

After the training of auto-encoder layers, the network fine-tuning for each modality is

performed by adding a softmax layer which predicts the final class. The intermediate features are the output of the fine-tuned network after removing the softmax layer. The hyper-parameters in the model such as the layer sizes, dropout parameters, and regularization coefficients are optimized using 10-fold cross-validation.

Intermediate Features for Imaging Data

First, we select the regions of interest and put them into a separate 3-dimensional convolutional neural network (Fig A2. in the supplementary material) with their weights shared across the CNN modules. CNN modules can extract higher level features from the abstraction of images to form concepts, that often correlate better with the targets. Each 3D CNN in the architecture above comprises 10 3D-convolutional kernels of size $5 \times 5 \times 5$ followed by pooling layers with pooling kernels of size $3 \times 3 \times 3$. After the pooling layer, we feed the pooled 3D images into Rectified Linear Unit (ReLU) non-linearities to learn complex features from the input modalities. We use volumetric batch normalization [312] that is an effective regularizer for convolutional neural networks. Next, the feature maps generated by each 3D CNN are flattened and fed into separate fully connected layers with ReLU activation functions, followed by drop-out regularizers. We integrate the features generated from each modality and feed them into the second level fully connected layer and the corresponding drop-out layer. Finally, we use a softmax layer with a negative-log-likelihood loss function to train the imaging network.

We use the combined features generated from the first level fully connected layers as the intermediate features that are fed into our multi-modality DL models.

6.2.4 Multimodal Data Integration

Data integration across modalities is increasingly being proposed as a method for bridging the gaps in our understanding of disease processes, and for improving clinical outcome predictions and the model performance. The integration of the data from different modalities can be performed at multiple levels (raw feature level, intermediate feature level, and decision level) and can follow different approaches during integration (concatenation- based integration, transformation based integration, and model-based integration)[313] (Figure 6.2). In this study, we propose an integration of the intermediate features (transformation based integration) generated in the previous step using a concatenation layer followed by a classification layer to get the Alzheimer's stage (Figure 6.3.). We tried k-nearest neighbors, decision trees, and support vectors machines as alternatives for the classification layer. In the event any modality was missing for a specific

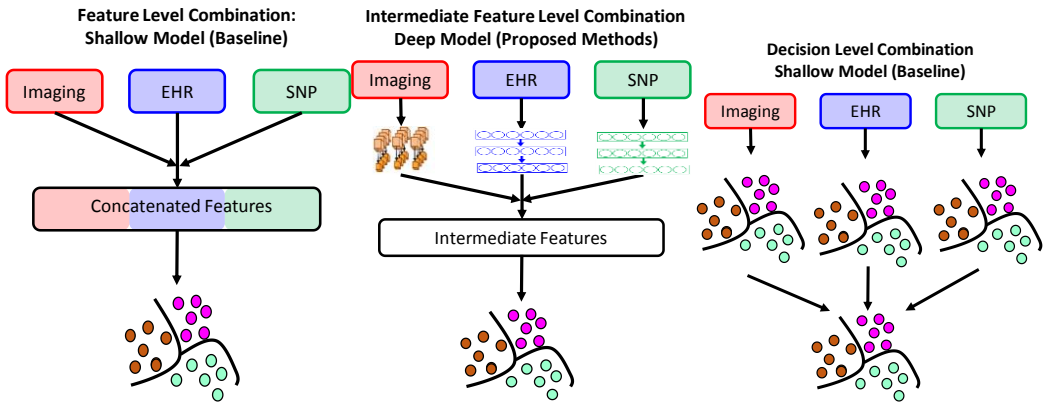


Figure 6.2: Deep Model for Data Integration Compared with Shallow Models of Data Integration. a) Feature level integration on shallow models, where the features are concatenated before passing into shallow models. b) Deep intermediate feature level integration where the original features are transformed separately using deep models prior to integration and prediction. c) Decision level integration where voting is performed using decisions of individual classifiers. In this study, we compare the performance of deep intermediate level integration against shallow feature and decision levels integrations for the prediction of Alzheimer's stages.

patient, we masked the modality with zeros. We evaluated our models using feature level combinations and decision level combinations as the baseline models.

6.2.5 Model Implementation

We developed the deep-models on torch lua environment. The libraries used included nn, nnx, optim, and autograd. We ran the models on the pace clusters service by Georgia Tech (https://pace.gatech.edu/overview). Pace clusters can be accessed by writing a short proposal detailing the project and its usage by PIs. Each new student account also requires a description of the student's project, GT ID and a mandatory tutorial. Due to a lower number of GPUs as compared to the CPU, we trained our models on CPUs. We used

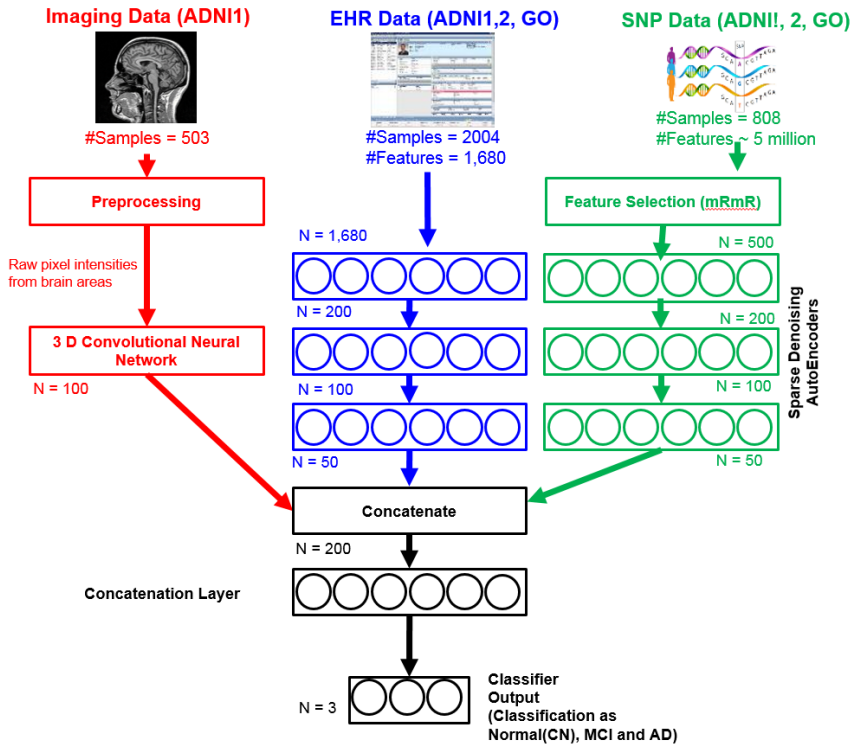


Figure 6.3: Intermediate-Feature-Level Combination Deep Models for Multimodality Data Integration for Clinical Decision Support. Data from diverse sources, imaging, EHR and SNP are combined using novel deep architectures. 3D convolutional neural network architectures used on 3D MR image regions to obtain intermediate imaging features. Deep stacked denoising autoencoders are used to obtain intermediate EHR features. Deep stacked denoising autoencoders are used obtain intermediate SNP features. The 3 types of intermediate features are passed into a classification layer for classification into Alzheimer's stages (CN, MCI and AD).

11 CPUs for training the deep models with each epoch taking about 1 min. For the baselines, and classification layers, we used matlab2016b also in pace clusters.

6.2.6 Model Evaluation

We first removed 10% of the data as an external test set. On the remaining 90%, we performed 10-fold cross-validation to optimize our models (hyper-parameters) as well as the baselines. For the integrated models, we evaluated the combination models. We test the integration models against feature and decision level combination as baselines. While testing individual classifiers, we use k-nearest neighbors (kNN), one-vs-one coding SVM, random forests, and decision trees as baselines. For each of the models and the baselines, we report mean values of accuracy, precision, recall, and meanF1 scores.

6.3. Results & Discussion

As mentioned above, we demonstrated our results using ADNI dataset to show the improvement in the prediction when using deep models for individual modalities and the improvements gained from data integration.

6.3.1 3D Convolutional Neural Network (DL) is Superior to Shallow Models on Imaging MRI Data

One patient's imaging data consists of 9108 3D voxels of dimension 22×23×18, corresponding to each of the 5 selected brain areas. The number of nodes in DL models for the first-level fully connected layers = 5×20 = 100 and the number of nodes for the second level fully connected layer is 20. The results (Table 6.3a.) indicate that the CNN based imaging models outperform shallow models and give the best precision and meanF1 scores.

6.3.2 Deep Autoencoder Model is Comparable to Shallow Models on EHR Data

EHR data consists of 2,004 patients with 1,680 normalized features per patient,

which we use to classify the patients into AD, MCI, and controls (three class). We use a three-layer auto-encoder with 200, 100 and 50 nodes each. The training of the DL networks is done using Adam with a max epoch count (repetition of DL network training on the entire dataset to allow adequate training) of 25. After hyperparameter optimization, the regularization coefficients for initial training is fixed at 0.03 and those for fine tuning at 0.03. The dropout probability is set to 0.6 for all the layers. The results (Table 6.3b.) indicate that the autoencoders outperform shallow models such as kNN and SVM, and they are comparable to decision trees and random forests.

6.3.3 Deep Autoencoder Model is Superior to Shallow Models for SNP Data

Processed SNP data consists of 808 patients with 500 features (each with levels 1,2,3), which we use to classify the patients into AD/MCI vs controls (two class). Auto-encoder network consists of three hidden layers with 200, 100 and 50 nodes each. Using Adam optimization and a max epoch count of 30, the best performing models have regularization coefficients for initial training as 0.03 and those for fine tuning at 0.06. The corruption (dropouts) is 0.6 for each layer. The results (Table 6.3c.) indicate that the auto-encoder models outperform all the baselines models.

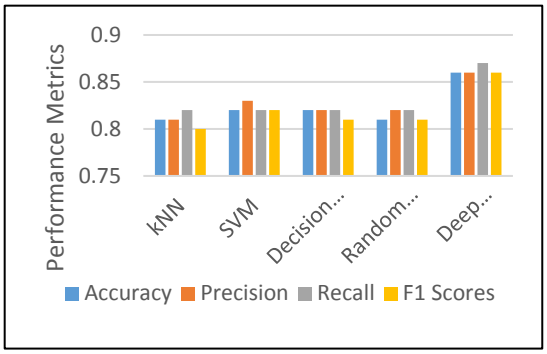
6.3.4 Results for Multi-Modality Classification

The intermediate features generated from the single-modality deep-models are concatenated and passed to an additional classification layer for integration.

Combination of all 3 modalities: (Imaging + EHR + SNP): Deep Model Outperforms

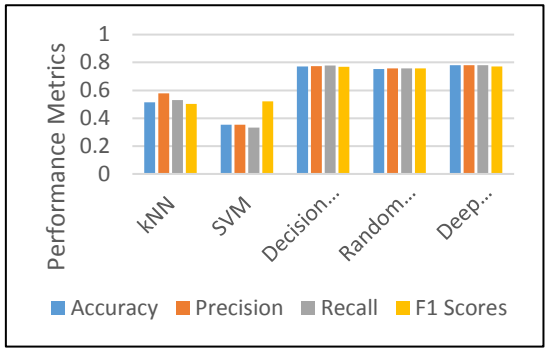
Table 6.3:: Internal Cross Validation Results for Individual Data Modality to Predict Alzheimer's Stage a) Imaging Results: Deep learning prediction performs better than shallow learning predictions b) EHR Results: Deep learning outperforms shallow models kNN and SVM and is comparable to decision trees and random forests c) SNP Results: Deep learning outperforms shallow models. The kNN, SVM , RF and decision trees are shallow models. ((kNN: k-Nearest Neighbors, SVM: Support Vector Machines, and RF: Random Forests).

Metrics		kNN	SVM	Decision Trees	RF	Deep Model
Accuracy	CN vs AD	0.81 ± 0.05	0.82 ± 0.09	0.82 ± 0.06	0.81 ± 0.08	**0.86 ± 0.04**
Precision	CN	0.78 ±0.12	0.82 ±0.09	0.82 ±0.14	0.81 ±0.1	**0.92 ±0.08**
	AD	**0.85 ±0.11**	0.84 ±0.13	0.82 ±0.1	0.82 ±0.13	0.80 ±0.1
Recall	CN	0.83 ±0.14	0.81 ±0.15	0.79 ±0.11	0.80 ±0.14	**0.85 ±0.08**
	AD	0.80 ±0.1	0.84 ±0.12	0.85 ±0.09	0.84 ±0.11	**0.89 ±0.1**
MeanF1	CN	0.79 ±0.06	0.80 ±0.1	0.79 ±0.08	0.80 ±0.07	**0.88 ±0.04**
	AD	0.81 ±0.05	0.83 ±0.1	0.83 ±0.06	0.82 ±0.09	**0.84 ±0.07**



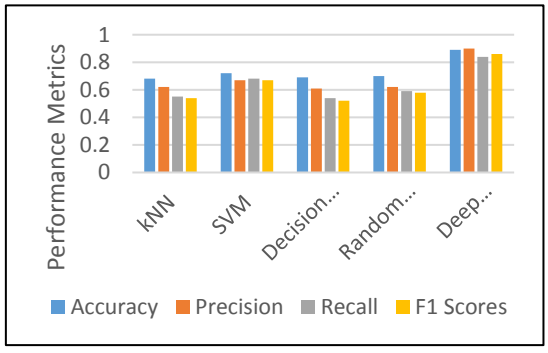
a: Alzheimer disease stage prediction using imaging modalities to predict CN vs AD. Convolutional neural networks (DL) outperformed all shallow models.

		kNN	SVM	Decision Trees	RF	Deep Model
Accuracy	CN	0.67 ± 0.04	0.84 ± 0.03	**0.9 ± 0.02**	0.88 ± 0.03	0.83 ± 0.07
	MCI	0.65 ± 0.03	0.73 ± 0.02	**0.79 ± 0.02**	0.76 ± 0.04	0.74 ± 0.06
	AD	0.78 ± 0.04	0.81 ± 0.02	0.82 ± 0.01	0.83 ± 0.03	**0.85 ± 0.03**
Precision	CN	0.51 ±0.03	0.77 ±0.04	**0.84 ±0.04**	0.81 ±0.05	0.75 ±0.12
	MCI	0.56 ±0.06	0.61 ±0.03	**0.76 ±0.02**	0.67 ±0.05	0.65 ±0.09
	AD	**0.86 ±0.07**	0.77 ±0.03	0.73 ±0.02	0.79 ±0.05	0.84 ±0.07
Recall	CN	0.88 ±0.09	0.77 ±0.08	**0.91 ±0.03**	0.84 ±0.05	0.76 ±0.27
	MCI	0.36 ±0.07	0.64 ±0.05	0.58 ±0.07	**0.66 ±0.07**	0.65 ±0.12
	AD	0.61 ±0.08	0.74 ±0.05	**0.84 ±0.05**	0.77 ±0.05	0.79 ±0.05
MeanF1	CN	0.64 ±0.03	0.77 ±0.05	**0.87 ±0.03**	0.82 ±0.04	0.72 ±0.23
	MCI	0.44 ±0.06	0.62 ±0.03	**0.66 ±0.05**	0.66 ±0.05	0.64 ±0.05
	AD	0.71 ±0.06	0.75 ±0.03	0.78 ±0.02	0.78 ±0.04	**0.81 ±0.04**



b: Alzheimer disease stage prediction using EHR modalities to predict CN vs MCI vs AD. Auto encoder networks (DL) outperformed shallow models kNN and SVM and was comparable to decision trees and RF.

		kNN	SVM	Decision Trees	RF	Deep Model
Accuracy	CN vs AD/MCI	0.68 ± 0.04	0.72 ± 0.07	0.69 ± 0.06	0.7 ± 0.04	**0.89 ± 0.03**
Precision	CN	0.51 ±0.27	0.53 ±0.1	0.50 ±0.2	0.48 ±0.13	**0.90 ±0.11**
	AD/MCI	0.73 ±0.04	0.81 ±0.06	0.73 ±0.03	0.75 ±0.05	**0.89 ±0.05**
Recall	CN	0.24 ±0.09	0.57 ±0.1	0.17 ±0.09	0.31 ±0.15	**0.72 ±0.11**
	AD/MCI	0.87 ±0.08	0.78 ±0.07	0.91 ±0.1	0.87 ±0.05	**0.96 ±0.05**
MeanF1	CN	0.29 ±0.08	0.54 ±0.09	0.24 ±0.1	0.36 ±0.13	**0.79 ±0.05**
	AD/MCI	0.79 ±0.03	0.79 ±0.05	0.80 ±0.05	0.80 ±0.03	**0.92 ±0.02**



c: Alzheimer disease stage prediction using SNP modalities to predict CN vs MCI/AD. Auto encoder networks (DL) outperformed all shallow models.

Shallow Models.

When a particular modality is not available, we mask it as zeros when using DL. The intermediate features from the three modalities are passed to the classification layer. We test kNN, decision trees, random forests, and support vectors machines as alternatives for the classification layer. Internal cross-validation (CV) accuracy (Table 6.4a) using deep models followed by random forests as the classification layer are the best. Deep models for the combination of the three modalities outperform single-modalities DL. In addition, during combination deep model outperforms shallow models such as feature-level and decision-level for both CV and external test sets (Table 6.5.).

Combination of SNP and EHR modalities: Deep Model Outperforms Shallow Models.

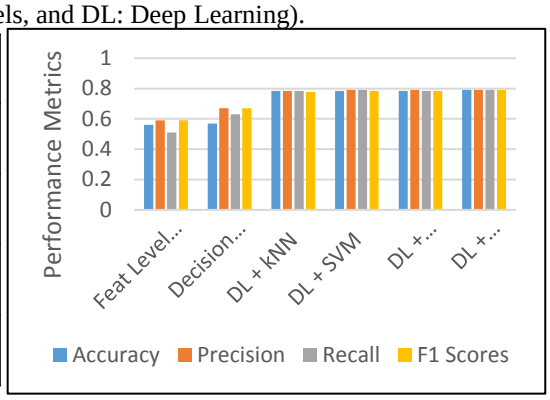
Internal CV accuracy of 0.78 ± 0 using deep models followed by random forests as the classification layer (Table 6,4b.) are the best. The deep models for EHR + SNP combinations outperform single-modalities DL. During combination, deep model outperforms shallow models such as feature-level combination models for both CV and external test sets (Table 6.5.).

Combination of Imaging and EHR modalities: Deep Model Outperforms Shallow Models.

Internal CV accuracy of 0.79 ± 0 using deep models followed by random forests and SVM as the classification layers (Table 6.4c.) are the best. The deep models for EHR + imaging combinations outperform single-modalities DL. In addition, during combination, DL model outperforms shallow models such as feature decision-level combination models for both CV and external test sets (Table 6.5.). Random forests as the

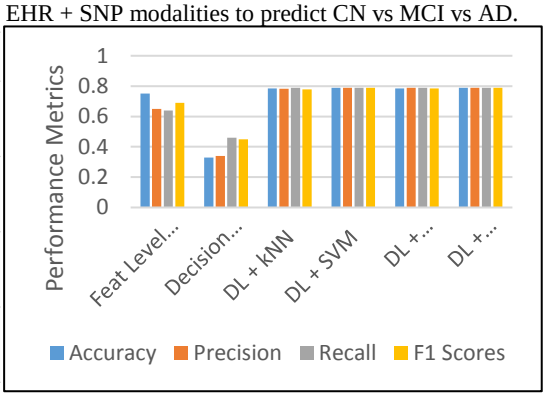
Table 6.4: Internal Cross Validation Results for Integration of Data Modalities to Predict Alzheimer's Stage a, b, c) Deep learning prediction performs better than shallow learning predictions b) Deep learning prediction performs better than shallow learning predictions d) Shallow learning gave a better prediction than deep learning due to small sample sizes. (kNN: k-Nearest Neighbors, SVM: Support Vector Machines, RF: Random Forests, SM: Shallow Models, and DL: Deep Learning).

		Feature Level (SM)	Decision Level (SM)	DL + kNN	DL + SVM	DL + Decision Trees	DL + RF
Accuracy	CN	0.73 ± 0.13	0.67 ± 0.13	0.87 ± 0.02	0.87 ± 0.02	0.87 ± 0.03	**0.88 ± 0.02**
	MCI	0.57 ± 0.12	0.7 ± 0.12	0.79 ± 0.03	0.79 ± 0.02	0.79 ± 0.04	**0.8 ± 0.02**
	AD	0.61 ± 0.13	0.64 ± 0.09	0.87 ± 0.02	**0.87 ± 0.02**	0.87 ± 0.03	**0.87 ± 0.02**
Precision	CN	0.64 ±0.16	0.51 ±0.17	0.76 ±0.05	0.79 ±0.02	0.79 ±0.05	**0.81 ±0.05**
	MCI	0.24 ±0.17	0.56 ±0.2	0.72 ±0.05	0.70 ±0.04	0.70 ±0.06	**0.72 ±0.03**
	AD	0.62 ±0.14	**1 ±0**	0.87 ±0.04	0.87 ±0.04	0.87 ±0.05	0.86 ±0.04
Recall	CN	0.70 ±0.21	**0.93 ±0.09**	0.9 ±0.04	0.84 ±0.05	0.85 ±0.06	0.85 ±0.06
	MCI	0.22 ±0.16	0.70 ±0.2	0.68 ±0.07	**0.71 ±0.05**	0.71 ±0.08	0.71 ±0.07
	AD	0.62 ±0.24	0.27 ±0.17	0.78 ±0.06	0.8 ±0.03	0.79 ±0.04	**0.81 ±0.05**
MeanF1	CN	0.66 ±0.16	0.65 ±0.14	0.82 ±0.03	0.81 ±0.03	0.82 ±0.04	**0.83 ±0.04**
	MCI	0.26 ±0.09	0.6 ±0.14	0.69 ±0.05	0.70 ±0.04	0.70 ±0.05	**0.71 ±0.03**
	AD	0.61 ±0.17	0.45 ±0.17	0.82 ±0.03	0.83 ±0.02	0.82 ±0.03	**0.83 ±0.03**



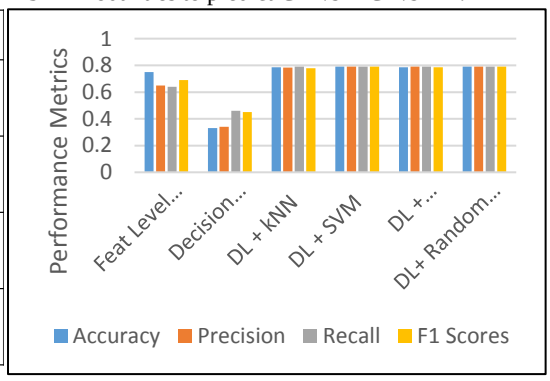
a: Alzheimer disease stage prediction using Imaging + EHR + SNP modalities to predict CN vs MCI vs AD.

		Feature Level (SM)	Decision Level (SM)	DL + kNN	DL + SVM	DL + Decision Trees	DL + RF
Accuracy	CN	0.87 ± 0.03	0.77 ± 0.04	**0.88 ± 0.02**	**0.88 ± 0.02**	0.87 ± 0.03	**0.88 ± 0.02**
	MCI	0.77 ± 0.06	0.76 ± 0.06	0.79 ± 0.03	**0.79 ± 0.01**	0.79 ± 0.04	0.79 ± 0.02
	AD	0.82 ± 0.04	0.78 ± 0.04	0.86 ± 0.02	**0.87 ± 0.01**	0.87 ± 0.03	0.87 ± 0.02
Precision	CN	**0.81 ±0.08**	0.59 ±0.05	0.79 ±0.03	0.79 ±0.04	0.79 ±0.05	0.79 ±0.04
	MCI	0.72 ±0.07	0.74 ±0.08	**0.72 ±0.06**	0.71 ±0.03	0.7 ±0.06	0.71 ±0.04
	AD	0.72 ±0.09	**1 ±0**	0.85 ±0.05	0.87 ±0.03	0.87 ±0.05	0.85 ±0.03
Recall	CN	0.78 ±0.06	**1 ±0**	0.87 ±0.05	0.88 ±0.04	0.85 ±0.07	0.87 ±0.03
	MCI	0.75 ±0.1	0.74 ±0.06	0.68 ±0.08	0.7 ±0.06	**0.71 ±0.07**	0.69 ±0.06
	AD	0.7 ±0.1	0.31 ±0.12	0.8 ±0.07	**0.79 ±0.04**	0.79 ±0.05	0.8 ±0.05
MeanF1	CN	0.8 ±0.06	0.74 ±0.04	0.83 ±0.03	**0.83 ±0.02**	0.82 ±0.05	0.83 ±0.03
	MCI	0.73 ±0.07	**0.74 ±0.06**	0.69 ±0.04	0.7 ±0.03	0.7 ±0.05	0.7 ±0.04
	AD	0.71 ±0.07	0.46 ±0.15	0.82 ±0.04	**0.83 ±0.02**	0.83 ±0.03	0.83 ±0.03



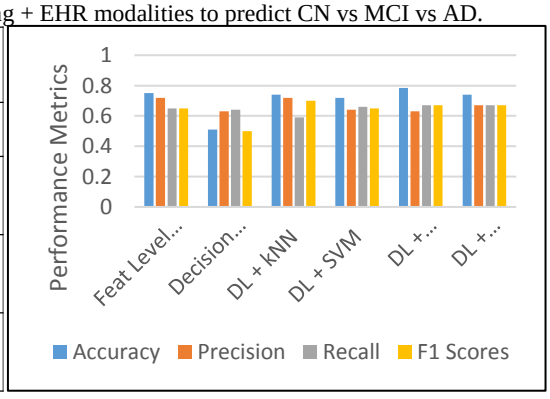
b: Alzheimer disease stage prediction using EHR + SNP modalities to predict CN vs MCI vs AD.

		Feature Level (SM)	Decision Level (SM)	DL + kNN	DL + SVM	DL + Decision Trees	DL + RF
Accuracy	CN	0.85 ± 0.05	0.58 ± 0.09	0.86 ± 0.03	**0.88 ± 0.03**	**0.88 ± 0.03**	0.88 ± 0.04
	MCI	0.78 ± 0.05	0.38 ± 0.1	**0.8 ± 0.03**	**0.8 ± 0.03**	0.79 ± 0.04	**0.8 ± 0.03**
	AD	0.83 ± 0.08	0.38 ± 0.08	0.87 ± 0.03	**0.88 ± 0.02**	0.87 ± 0.03	0.87 ± 0.03
Precision	CN	0.7 ±0.06	0.52 ±0.08	0.75 ±0.04	**0.8 ±0.06**	**0.8 ±0.06**	0.79 ±0.04
	MCI	0.44 ±0.38	0.16 ±0.12	**0.74 ±0.05**	0.71 ±0.05	0.71 ±0.06	0.71 ±0.05
	AD	0.82 ±0.08	0 ±0	0.84 ±0.05	**0.87 ±0.04**	0.86 ±0.06	**0.87 ±0.04**
Recall	CN	0.87 ±0.07	**1 ±0**	0.85 ±0.04	0.86 ±0.04	0.85 ±0.07	0.87 ±0.08
	MCI	0.13 ±0.11	0.38 ±0.19	0.68 ±0.04	**0.71 ±0.08**	**0.71 ±0.08**	0.7 ±0.05
	AD	**0.92 ±0.07**	0 ±0	0.81 ±0.05	0.81 ±0.05	0.8 ±0.05	0.8 ±0.04
MeanF1	CN	0.77 ±0.06	0.68 ±0.07	0.8 ±0.04	0.82 ±0.04	0.82 ±0.05	**0.83 ±0.06**
	MCI	0.27 ±0.1	0.22 ±0.14	**0.71 ±0.04**	0.71 ±0.05	0.71 ±0.06	**0.71 ±0.04**
	AD	**0.86 ±0.07**	0 ±0	0.83 ±0.05	0.84 ±0.03	0.83 ±0.05	0.83 ±0.04



c: Alzheimer disease stage prediction using Imaging + EHR modalities to predict CN vs MCI vs AD.

		Feature Level (SM)	Decision Level (SM)	DL + kNN	DL + SVM	DL + Decision Trees	DL + RF
Accuracy	CN vs AD/MCI	**0.75 ± 0.11**	0.51 ± 0.12	0.74 ± 0.08	0.72 ± 0.06	0.73 ± 0.08	0.74 ± 0.09
Precision	CN	0.67 ±0.25	0.37 ±0.12	**0.74 ±0.08**	0.72 ±0.06	0.73 ±0.08	0.74 ±0.09
	AD/MCI	0.78 ±0.12	**0.9 ±0.13**	0.69 ±0.4	0.55 ±0.11	0.54 ±0.12	0.61 ±0.2
Recall	CN	0.39 ±0.15	**0.93 ±0.09**	0.74 ±0.07	0.79 ±0.06	0.81 ±0.05	0.8 ±0.06
	AD/MCI	**0.91 ±0.07**	0.34 ±0.15	0.22 ±0.24	0.49 ±0.18	0.54 ±0.16	0.5 ±0.16
MeanF1	CN	0.48 ±0.15	0.51 ±0.12	**0.97 ±0.05**	0.82 ±0.09	0.8 ±0.09	0.84 ±0.13
	AD/MCI	**0.83 ±0.09**	0.49 ±0.16	0.56 ±0.14	0.5 ±0.12	0.54 ±0.14	0.53 ±0.14



d: Alzheimer disease stage prediction using Imaging + SNP modalities to predict CN vs MCI/AD.

classification layer give the best performance on the external set.

Combination of Imaging and SNP modalities: Shallow Model Outperforms Deep Models.

We perform two-class classification using a combination of SNP and imaging intermediate features (CN vs AD/MCI). Internal CV accuracy of 0.75 ± 0.11, using feature-level combination models (Table 6.4d) is the best. However, the results on the external data are poor. This can be attributed to the small overlap of 220 samples between the two

Table 6.5: Features extraction from deep models and comparison of internal validation results with external test result. Autoencoder models are preferred for EHR and SNP data and CNN for imaging data. For multi-modality models, the three modality models and two modality models (EHR + SNP, EHR + imaging gave the best prediction performance). For the multi-modality models, 3 or 4 combinations deep models outperformed shallow models.

	Models	**Internal Cross Validation Performance**	**External Test Performance**
EHR (Deep Models) (CN,MCI,AD)	Regularization coefficients (.03,.03) Dropouts (0.6,0.6,0.6) Layer sizes (200,100,75)	Accuracy: 0.78 ± 0.03 Precision: 0.78 ± 0.04 Recall: 0.78 ± 0.05 F1 Scores: 0.77 ± 0.04	Accuracy: 0.76 Precision: 0.76 Recall: 0.77 F1 Scores: 0.76
Imaging (Deep Models) Prediction (CN,AD)	Highest on validation (Dropout- .5, Batch size 5 , Layer size(20), # areas = 5) Highest on external test (SVM kernel = linear)	Accuracy: 0.86 ± 0.04 Precision: 0.86 ± 0.04 Recall: 0.87 ± 0.04 F1 Scores: 0.86 ± 0.04	Accuracy: 0.84 Precision: 0.83 Recall: 0.83 F1 Scores: 0.83
SNP (Deep Models) Prediction (CN,MCI/AD)	Regularization coefficients (.03,.03), Dropouts(0.6,0.6,0.6) Layer sizes(200,100,50)	Accuracy: 0.89 ± 0.03 Precision: 0.9 ± 0.04 Recall: 0.84 ± 0.03 F1 Scores: 0.86 ± 0.04	Accuracy: 0.66 Precision: 0.66 Recall: 0.57 F1 Scores: 0.53
EHR + SNP + Imaging (Deep Models) Prediction (CN,MCI,AD)	Regularization coefficients (.03,.03) Dropouts(0.6,0.6,0.6) Layer sizes(200,100,50) Random Forest Trees = 31	Accuracy: 0.79 ± 0 Precision: 0.79 ± 0.07 Recall: 0.79 ± 0.07 F1 Scores: 0.79 ± 0.07	Accuracy: 0.78 Precision: 0.77 Recall: 0.78 F1 Scores: 0.78
EHR + SNP (Deep Models) Prediction (CN,MCI,AD)	Regularization coefficients (.03,.03) Dropouts(0.6,0.6,0.6) Layer sizes(200,100,50) Random Forest Trees = 31	Accuracy: 0.78 ± 0 Precision: 0.79 ± 0.07 Recall: 0.79 ± 0.09 F1 Scores: 0.79 ± 0.07	Accuracy: 0.78 Precision: 0.78 Recall: 0.79 F1 Scores: 0.78
EHR + Imaging (Deep Models) Prediction (CN,MCI,AD)	Regularization coefficients (.03,.03) Dropouts(0.6,0.6,0.6) Layer sizes(200,100,50) Random Forest Trees = 31;	Accuracy: 0.79 ± 0 Precision: 0.79 ± 0.08 Recall: 0.79 ± 0.08 F1 Scores: 0.79 ± 0.07	Accuracy: 0.77 Precision: 0.76 Recall: 0.77 F1 Scores: 0.77
SNP + Imaging (Shallow Models) Prediction (CN,MCI/AD)	Random Forest Trees = 20;	Accuracy: 0.75 ± 0.11 Precision: 0.72 ± 0.16 Recall: 0.65 ± 0.09 F1 Scores: 0.65 ± 0.12	Accuracy: 0.63 Precision: 0.62 Recall: 0.57 F1 Scores: 0.56

modalities.

6.3.5 Discussion for Novel DL and Multi-Modality Data Analysis

Our results indicate that the deep models outperform traditional shallow models for single-modalities. This is because shallow models require expert crafted features that are mutually uncorrelated, while deep models can find the optimal set of features during training. In addition, deep models such as autoencoders and CNNs perform unsupervised feature generation, which allows us to combine the models with more sophisticated decision layer and facilitates the modeling of complex decision boundaries for multiclass class problems [314]. Due to this property, deep models are particularly effective for the identification of MCI, which has been a clinical challenge in Alzheimer's research due to small differences between the three groups. Shallow models (except random forests) also do not tolerate noisy and missing data or missing modalities well. As a result, in noisy data, DL gives the best performance for single-modalities.

Integration of multiple modalities improves the prediction accuracy (three of four scenarios). The deep models for integration also show improved performance over traditional feature-level and decision-level integrations. The superior performance of the DL is due to its ability to extract relationships amongst features from different modalities. When the dataset is very small (combination of Imaging and SNP), deep models do not perform well. This is due to the fact that deep models need larger datasets for training [315]. Overall our investigations show that

- for single-modality data (clinical, and imaging), the performance of DL models are always better than those of shallow models; and
- when using DL models, predictions by multi-modality data is better than

that of single-modality data, and the three best fusion set ups are: EHR + SNP, EHR + Imaging + SNP, and EHR + Imaging.

6.3.6 Interpretation of Deep-Models:

Model interpretation is a major challenge for deep learning, which is often considered as a barrier for real world applications of deep models in the biomedical domain. Research has shown that weights of deep models affect the results through several layers and combinations, hence do not yield clinically meaningful interpretations [235].

As mentioned above, interpretability of deep-learning models is challenging.

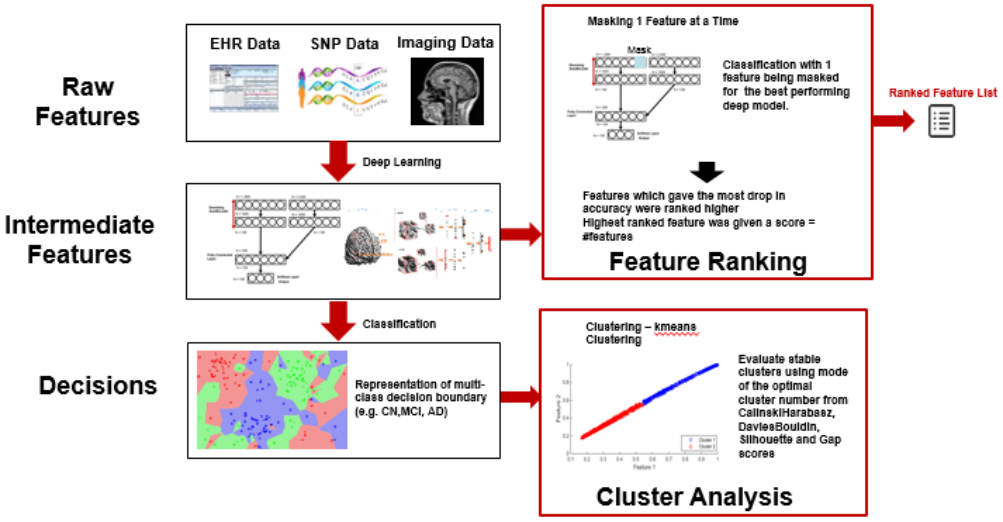


Figure 6.4: Model interpretation pipeline. The features for the deep models are masked one at a time and the effect on the classification is observed. The feature which gives the highest drop in accuracy is ranked the highest. Once we ranked the features, we checked if the intermediate picked associations different from raw data using cluster analysis.

However, the Interpretability of deep-learning models constitutes a major challenge. In this work, we interpret the models by masking one feature at a time, the features which gave highest drop in accuracy was picked as the top features. For each repeat, we gave the top features, the highest score. We rank the feature with the highest average score the top

Table 6.6: Top 10 features for AD classification from masking features

SN	EHR	Imp 90% features = 1509	SNP		Imp 90% features = 423	IM	Imp 90% features = 5	EHR+SNP	Imp 90% features = 1754	EHR+IM	Imp 90% features = 1511	EHR+SNP+IM	Imp 90% features = 520	SNP+IM		Imp 90% features = 29
1	Clinical Dementia Rating Communication	0.29033 1	4	7537597	0.066667	Right hippocampus	0.3237299	Clinical Dementia Rating Judgement and Problem Solving Score	0.065255	Executive function summary score	0.054174	Clinical Dementia Rating Memory	0.059638	7	103395766	0.4544431
2	Family History Question (Who answered)	0.29033 1	10	6885 1845	0.052381	Left hippocampus	0.3152544	Alzheimer's Disease Assessment Scale 13	0.065255	Memory summary score	0.054174	Clinical Dementia Rating Global	0.04858 8	Left AmygdalaVox sum		0.4421544
3	Cortical Thickness Average of Right Parahippocampal (Imaging Summary)	0.29033 1	10	6884 3778	0.052381	Right superior temporal	0.3152544	Alzheimer's Disease Assessment Scale 13 baseline	0.065255	Volume (aseg.stat) of LeftHippocampus	0.048649	Medical History Hematopoietic-Lymphatic Disorders	0.043063	Right hippocampus Vox mean		0.4177761
4	Cortical Thickness Average of Left Parahippocampal (Imaging Summary)	0.29033 1	10	6884 7107	0.052381	Left amygdala	0.3152544	Clinical Dementia Rating Global	0.05973	Mini Mental State Exam (Recall Date)	0.048649	Cortical Thickness Standard Deviation of LeftLateralOrbitofrontal	0.043063	6	151321980	0.3988485
5	Cortical Thickness Average of Right Caudal Middle Frontal (Imaging Summary)	0.29033 1	10	6884 6766	0.052381	Right amygdala	0.306788	Functional Assessment Questionnaire Remembering appointments, family occasions, holidays, medications.	0.05973	Mini Mental State Exam (Total Score)	0.048649	Segmented ventricular volume	0.043063	Right AmygdalaShannon Entropy		0.3873888

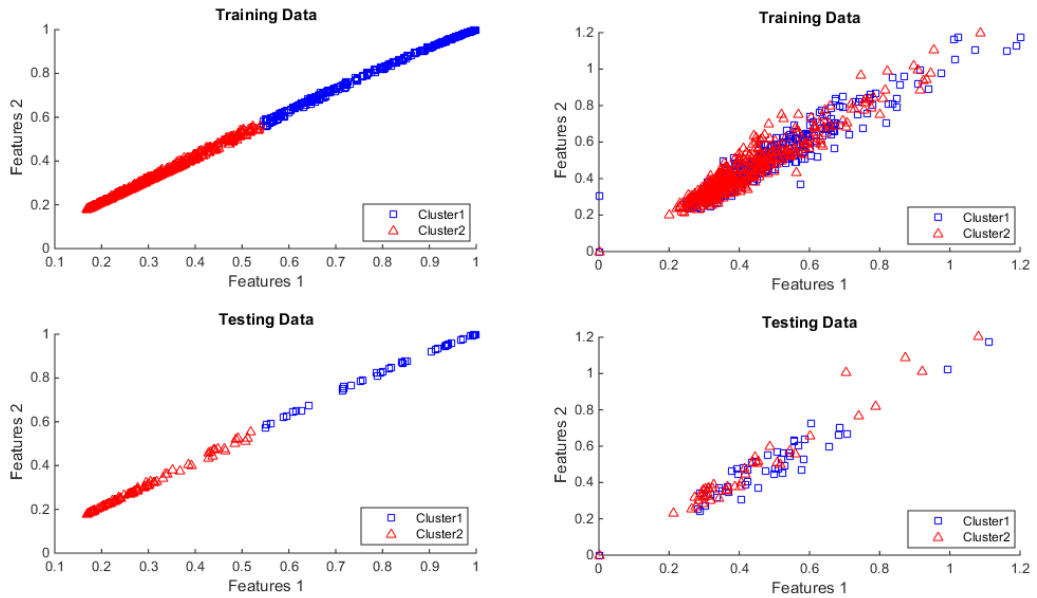
feature. For interpretation, we picked the model with the performance in each of the four integrations and for each individual model (Figure 6.4).

The top EHR features (Table 6.6) include memory tests, imaging summary scores, and brain volumes. Changes to memory and brain volumes have been reported as AD biomarkers. Imaging markers such as involvement of limbic and cortical regions [288],

and changes in hippocampus volume and structure [289, 316] are known biomarkers in positron emission tomography (PET) and magnetic resonance imaging (MRI) studies. SNP features picked chromosome 10, and 4.

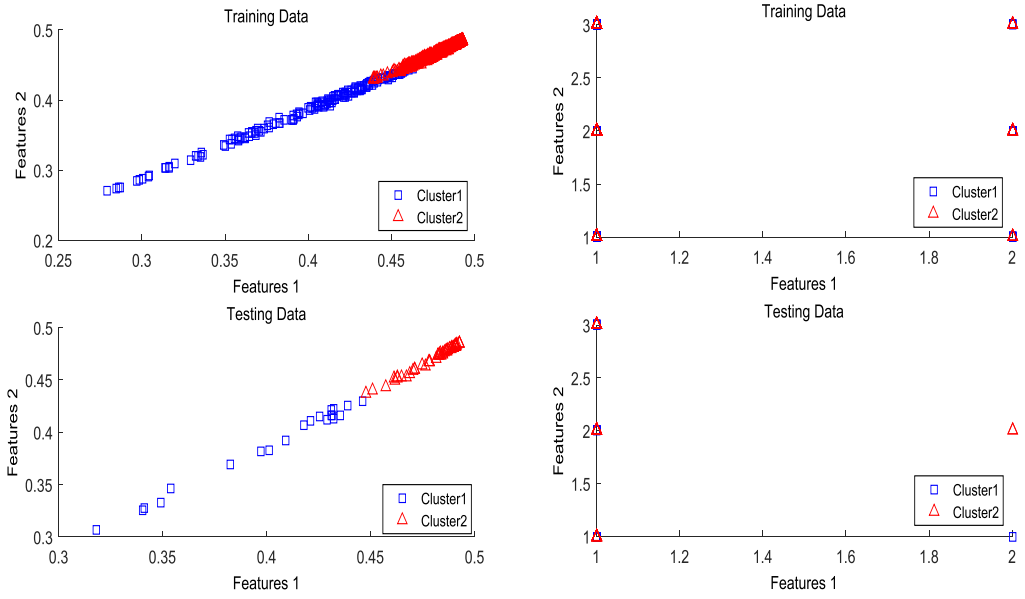
SNP + Imaging + EHR and SNP + EHR pick more EHR features (memory tests, metabolic markers and brain volume) which are known AD related features. EHR + Imaging pick EHR features including brain volumes, clinical dementia ratings, and metabolite markers. Imaging + SNP pick brain areas such as the hippocampus, and amygdala higher than SNP features.

In addition, we also clustered the intermediate features from EHR and SNP data were first clustered using kmeans to show associations in intermediate features. We tested



a b

Figure 6.5: Cluster analysis results: EHR Data. a) gives the intermediate features plotted against the top ranked features which discriminated the two clusters using relieFF. b) a) gives the original data plotted against the top ranked features which discriminated the two clusters using relieFF



a b

Figure 6.6: Cluster analysis results: SNP Data. a) gives the intermediate features plotted against the top ranked features which discriminated the two clusters using relieFF. b) a) gives the original data plotted against the top ranked features which discriminated the two clusters using relieFF

cluster number of 2-20 to pick the optimum cluster number using clustering scores. The cluster number was evaluated using the mode of cluster number generated using Calinski Harabasz [267], Davies-Bouldin [268], silhouette [269] and gap scores [270]. Then we checked the features which discriminated the clusters using relieFF feature selection algorithm. The top features which consistently discriminated the clusters across multiple folds and repeats in the training and test sets for both EHR and SNP data are reported. On plotting the clusters for intermediate and raw features, we found that the intermediate features generated better separation as compared to the original features. This indicates subtle relationships in intermediate features, which were picked by deep-models (Figures 6.5, 6.6.).

The top EHR features which discriminated the two clusters include imaging score

summaries and memory tests. More specifically, the features include limbic medial-temporal GM, unit recall test, left frontal operculum, left superior temporal gyrus, executive function summary, story recall memory test, clinical dementia rating, left posterior orbital gyrus, left superior frontal gyrus, and left superior occipital gyrus. The top SNP features which discriminated the two clusters include chromosome 19, 22, and 10. All these chromosomes are well-known genes which are associated with Alzheimer's disease. The top SNP locations involved include 45396144, 46619419, 45387596, 45410444, 68854980, 68854824, 45396219, 68838270, 50862870, 68821956.

6.4. Conclusion and Key Innovations

Less than 50% of the people with Alzheimer's are diagnosed accurately on the basis of their clinical symptoms. Current biomarkers for AD are based on a single data modality and the prediction accuracy also remains low for patient stage classification. In this study, we integrate multiple datasets using deep learning methods (stacked de-noising auto-encoders, and 3D- convolutional neural network), to achieve synergistic boosts in accuracy. We demonstrate a statistically significant improvement over baseline models including, support vector machines, decision trees, and k nearest neighbors on ADNI dataset. Despite the improved performance, our study suffers from short-comings such as limited dataset sizes . In the future, we will test our models on a larger and richer dataset.

To summarize, the key innovations of this chapter include:

- Deep-models outperform shallow models for single-modality Alzheimer's stage prediction.
- Novel DL framework for multiple-modality data fusion outperforms single-modality DL.

- Novel perturbation and clustering based feature extraction assisting DL model interpretations are capable of AD stage prediction.
- Application of 3D convolutional neural network architecture for MRI image data analysis in Alzheimer's disease.

Despite the improved performance, our study suffers from short-comings such as limited dataset sizes. In the future, we will test our models on a larger and richer dataset.

CHAPTER VII

CONCLUSION & FUTURE WORK

The concrete goals of this dissertation were to develop decision support tools for the prediction of adverse outcomes using electronic health records. The specific technical achievements of this dissertation corresponding to the three research objectives are (Figure 7.1):

1. Development and validation of quality control measures and novel imputation techniques for multiple types of missing data in EHR.
2. Construction of predictive models and visualizations using temporal predictive models applicable to adverse event detection such as ICU mortality, ICU readmission, sepsis and respiratory distress syndrome.

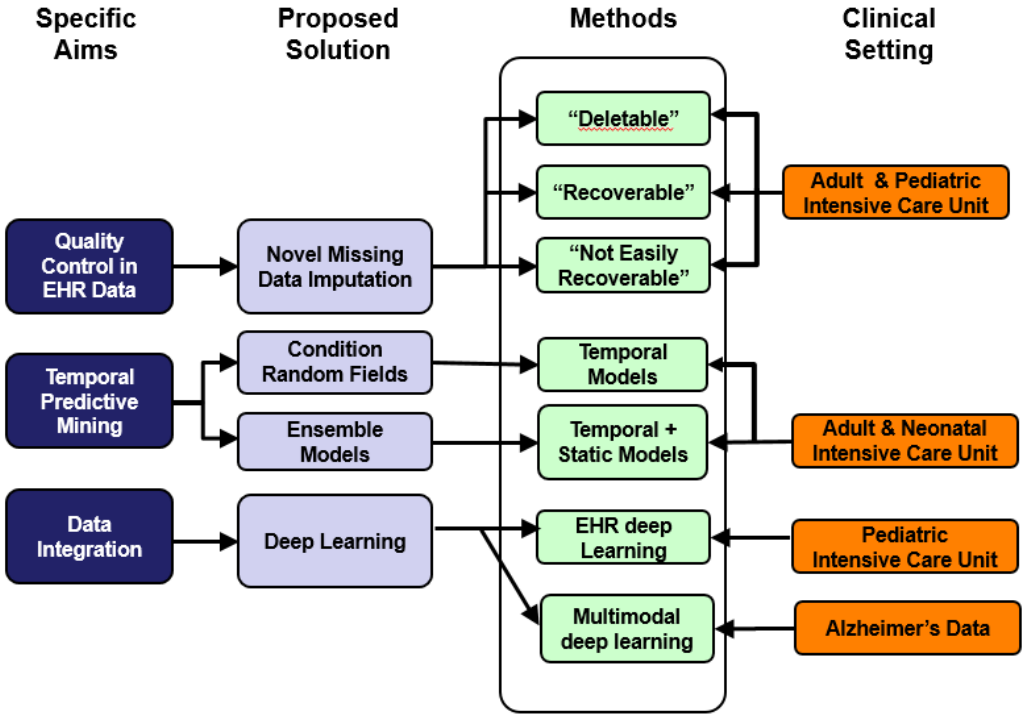


Figure 7.1: Summary of dissertation topic

3. Development and validation of integrated heterogenous temporal sequences and genetic data using deep learning for predicting adverse condition such as long ICU stay and Alzheimer's disease stages.

7.1. Concrete Innovation Deliverables

The key innovations of this dissertation, as noted at the closing of each chapter, are summarized below:

- (Chapter 2) Categorization of missing data in EHR into multiple types.
- (Chapter 2) Development of novel missing data types for multiple types of missing data.
- (Chapter 2) Evaluation of missing data imputation using adult and pediatric datasets.
- (Chapter 3) Development of time series data analysis models for ICU data without explicit independence assumptions.
- (Chapter 3) Analysis of adults and pediatric models for ICU mortality, 30 day ICU readmission, sepsis and RDS.
- (Chapter 3) First study to combine CRF with survival curves to show individual patient risk profiles
- (Chapter 4) Combination of static and temporal data for mortality and 30 day ICU readmission.
- (Chapter 4) Integration demonstrates synergistic improvement for integration models over individual modalities.
- (Chapter 5) Development of a framework for combining data from multiple temporal resolutions using deep models.

- (Chapter 5) Evaluation of integration models to demonstrate ICU length of stay prediction on pediatric ICU patient data.
- (Chapter 5) Development of a clustering based analysis for data interpretation.
- (Chapter 6) Development of a framework for combining data from multiple sources using deep models.
- (Chapter 6) Evaluation using Alzheimer's disease dataset to demonstrate that deep models outperformed shallow models within each modality and combination models outperformed individual models.
- (Chapter 6) Development of a clustering based analysis for data interpretation.

7.2. Concrete Publication Deliverables

The section provides a comprehensive list of publications completed during my years as a Ph.D. student.

Published or Accepted for Publication

Journals

- C. Cheng, N. Chanani, **J. Venugopalan**, K. Maher, and D. Wang, "icuARM–An ICU Clinical Decision Support System Using Association Rule Mining," 2013.
- P. Wu, C. Cheng, C Kaddi, **J Venugopalan**, R Hoffman, and MD Wang, " -Omic and Electronic Health Records, Big Data Analytics for Precision Medicine," in Transactions in Biomedical Engineering 2016.
- **J Venugopalan**, N. Chanani, K. Maher, and MD. Wang, " Novel Data Imputation for Multiple Types of Missing Data in Intensive Care Units," in Journal of Biomedical and Health Informatics, 2017.

- ***J Venugopalan,*** *Y Sha**,** T Buchman**,** MD Wang. "Data quality control in critical care." in SCCM Medicine* (Under Review).

Conferences

- ***J. Venugopalan****, C. Chihwen, and M. D. Wang, "MotionTalk: Personalized home rehabilitation system for assisting patients with impaired mobility," in ACM Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB) 2014 Annual International Conference of ACM 2014.*
- ***J. Venugopalan****, C. Chihwen, T. H. Stokes, and M. D. Wang, " Kinect-based Rehabilitation System for Patients with Traumatic Brain Injury," in Engineering in Medicine and Biology Society (EMBC), 2013 Annual International Conference of the IEEE, 2013.*
- ***J. Venugopalan****, C. Brown, C. Chihwen, T. H. Stokes, and M. D. Wang, "Activity and school attendance monitoring system for adolescents with Sickle cell disease," in Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE, 2012, pp. 2456-2459.*
- *J. Ren,* ***J. Venugopalan****, J. Xu, B. Kairdolf, and M. D. Wang, " Multi-Channel LED Light Source for Fluorescent Agent Aided Minimally Invasive Surgery", in Engineering in Medicine and Biology Society (EMBC), 2014 Annual International Conference of the IEEE, 2014.*
- *R. Durfee,* ***J. Venugopalan****, J. Ren , and M. D. Wang, " Multi-Channel LED Light Source for Fluorescent Agent Aided Minimally Invasive Surgery", in Point of Care (POC), 2014 Annual International Conference of the IEEE, 2014.*

- *C. Cheng, R. C. Brown, L. L. Cohen, **J. Venugopalan**, T. H. Stokes, and M. D. Wang, "iACT-An interactive mHealth monitoring system to enhance psychotherapy for adolescents with sickle cell disease," in Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE, 2013, pp. 2279-2282.*
- *TH Stokes, **J Venugopalan**, EN Hubbard, and MD Wang. "A pilot biomedical engineering course in rapid prototyping for mobile health. " Conf Proc IEEE Eng Med Biol Soc, EMBC. 2013 Jul 3; 2515-2518.*
- *I Raharjo, **J Venugopalan,** T Burns**,** MD Wang. "Development of user-friendly and interactive data collection system for cerebral palsy." Biomedical and Health Informatics (BHI), 2016.*
- *Y. Sha, **J Venugopalan**, N. Chanani, K. Maher, and M.D. Wang, " A Novel Temporal Similarity Measure for Patients Based on Irregularly Measured Data in Electronic Health Records," in ACM Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB) 2016 Annual International Conference of ACM 2016.*
- ***J Venugopalan**, M LaPlaca, and MD Wang, " Mining standardized neurological signs and symptoms data for concussion identification." in Biomedical and Health Informatics (BHI), 2017 IEEE-EMBS International Conference on, 2017.*
- ***J Venugopalan**, N Chanani, KO Maher, and MD. Wang. "Combination of static and temporal data analysis to predict mortality and readmission in intensive care unit" in " Conf Proc IEEE Eng Med Biol Soc, EMBC. 2017.*

Abstracts

- J. Venugopalan, R. Hoffman, C. Cheng, M. D. Wang, “Time-series data analysis to predict mortality and cardiac arrest in pediatric populations,” in 2014 Pediatric Healthcare Innovation Conference
- J. Venugopalan, C. Chihwen, T. H. Stokes, and M. D. Wang, "Cloud –Based Integrative Pain Management System"—Wireless Health 2013
- J. Venugopalan, C. Chihwen, T. H. Stokes, and M. D. Wang, " Kinect-based Rehabilitation System for Patients with Traumatic Brain Injury,"- BMES 2013
- Venugopalan, C. Brown, C. Chihwen, T. H. Stokes, and M. D. Wang, "Activity and school attendance monitoring system for adolescents with Sickle cell disease," in BMES 2012
- Venugopalan J, Chanani N, Maher KO and Wang MD. "Data quality control for the improved prediction of readmission in intensive care unit" in Annual Critical Care Congress,SCCM, 2015

Manuscripts under Review

- ***J Venugopalan**, N Chanani, KO Maher, and MD. Wang. “Time-series analysis of intensive care data to predict mortality and readmission.” in Journal of Biomedical Informatics **[***
- ***J Venugopalan,** H Hassanzadeh**,** L Tong**,** and MD Wang. “Multimodal analysis using deep-learning for Alzheimer's disease detection”, in Journal of American Medical Informatics Association.*
- ***J Venugopalan,** N Chanani, KO Maher**,** and MD Wang. “Integration of static and temporal data analysis using deep learning to predict length of stay in the intensive care unit ”*

7.3. Directions for Future Research and Concluding Remarks

The models and tools developed in this dissertation are complete and fully functional. However, it is essential to identify potential directions for future research to help improve the current work and applications. The specific potential research extensions are discussed at the end of each chapter. In addition to those, the high level extensions of this thesis work are discussed here. The opportunities are can be application novelties and data-mining opportunities.

7.3.1 Application Opportunities

In this chapters, I have showcased the application of data imputation, time-series analysis and data integration to datasets of varying sizes from a few hundred samples to a thousands of data samples. The ICU datasets included static data, temporal lab and vital signs data. We showcased prediction of ICU mortality, readmission, sepsis, RDS and length of stay for adult, pediatric and neonatal population. The techniques and tools developed in this dissertation can be readily deployed to higher frequency data such as waveform data. It can also be used in other research domains such as public health analysis, molecular data analysis, and insurance data analysis.

In addition, to this data sizes is a major challenge in healthcare research. Most data mining models, especially deep-learning models benefit greatly from larger datasets. Since data collection is expensive, creation of synthetic data using available patient data for training of the models has the potential for improving the prediction of the decision support systems. Generative models such as hidden markov models[317], generative adversarial networks[318] and generative long-short-term memory networks[319] are some candidate techniques which can be used for this analysis.

7.3.2 Data Mining Opportunities

Data Quality

Meaningful" use of EHR data and evidence-based medicine are concepts which have a great potential in the areas of comparative effectiveness research, predictive analytics, improved clinical outcomes research, personalized medicine, and precision medicine. However, clinically actionable results can be obtained only when the data is of sufficient quality to promote the decision support systems. In order to accelerate the quality of data in EHR and critical care systems, more research is needed in the areas of i) data normalization framework to integrate data from heterogeneous sources and multiple institutions (e.g. data ontology studies, HL7, FHIR[320], SHARPn [321], eMERGE [322], MIMIC [149]); ii) data quality indices development and evaluation (e.g. IBM Watson, SHARPn [321]; iii) data cleaning algorithms (e.g. error and artefact removal, data imputation); iv) natural language processing for clinical notes; and v) development of robust phenotyping algorithms (e.g. dealing with irregularly sampled data).

In addition, to effectively use EHR data, the different quality issues need to be addressed at different times in the healthcare process by the different stakeholders including investigators (e.g. researchers, statisticians, clinical PIs), users (e.g. physicians, business data analysts) , evaluators and policy makers.

Investigators

Quality control practices during data collection, aggregation, design of study experiments, and data analysis are very important for successful adoption of the results into clinical practice. Despite research, there is no clear consensus on the handling of the issues

in ICU data. Research is needed in the areas of data normalization from heterogeneous sources and multiple institutions; data quality indices and evaluation (e.g. IBM Watson, SHARPn [321]); data cleaning algorithms (e.g. error and artefact removal, data imputation); natural language processing; and the development of robust phenotyping algorithms (e.g. dealing with irregularly sampled data).

Users

Data quality at the time of data entry and interpretation of the studies are essential for improved quality of care. Several recommender systems for data entry and rules to detect outliers have been proposed and have been shown to have an impact on data quality. Data sharing, education, and involvement in the development of decision support have been proposed for improving data quality [323] and quality of care.

Evaluators and Policy Makers

The US Food and Drug Administration (FDA) is taking an active interest in clinical DSS, and mobile software [324]. The major concerns are addressed by FDA by creating specific regulations for the category wise division of adverse effects: (1) the acquisition of the wrong patient record or incorporation of misinformation, (2) the omission or obliteration of patient data, (3) falsified data processing, and (4) the incompatibility between multi-vendor application and systems [325, 326]. FDA is also monitoring decision support systems as a post-surveillance to minimize risk and to improve quality.

Temporal Data Mining

In this dissertation, we showcased the development of temporal data mining, and combination of heterogenous temporal resolution data using machine learning and deep-learning techniques. An extension of the current work would be to include irregularly sampled data for analysis. The few studies which deal with irregular sampling, primarily use imputation techniques and robust parameter extraction [327]. Some work on the visualization of irregularly sampled data to assist physicians to make a decision has also been done [110]. Most of this work is however concentrated on the waveform analysis. Research on the use of integrated clinical data with irregular sampling is still a more open research area.

Integration Research

In this dissertation, I showcase the integration of temporal and multiple different datatypes using deep-learning techniques. Due to limitation of dataset sizes, we chose relatively simple deep-learning models such as autoencoders, and LSTM networks. With the availability of large datasets, using techniques like transfer learning [328, 329] on large models developed using open source models developed for social network and imaging data are potential directions.

7.3.3 Concluding Remarks

In this dissertation, I have developed a suite of quality control, data mining and deep learning tools to address key challenges in clinical decision support research. It includes models for data imputation, time-series analysis and data integration that have been successfully applied to investigate adverse ICU outcomes. In the preceding sections, I have also discussed potentials for future investigations, building upon this work.

REFERENCES

[1] J. Venugopalan, N. Chanani, K. Maher, and M. D. Wang, "Novel Data Imputation for Multiple Types of Missing Data in Intensive Care Units," *Journal of Biomedical and Health Informatics*, 2016 (Under review).

[2] (2014). *National Health Expenditure Data*. Available: http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/index.html?redirect=/nationalhealthexpenddata/

[3] K. Davis, K. Stremikis, D. Squires, and C. Schoen, "Mirror, mirror on the wall," *How the performance of the US Health care system compares internationally. New York: CommonWealth Fund*, 2014.

[4] W. H. Organization. *"Global Health Expenditure Database"*. Available: http://apps.who.int/nha/database/ResourcesPage.aspx

[5] *Health statistics and information systems*. Available: http://www.who.int/healthinfo/en/

[6] H. Act, "Health Information Technology for Economic and Clinical Health," 2010.

[7] (Accessed Aug 08, 2013). *Patient Protection and Affordable Care Act (Wikipedia)*. Available: http://en.wikipedia.org/wiki/Patient_Protection_and_Affordable_Care_Act

[8] D. Blumenthal and M. Tavenner, "The "meaningful use" regulation for electronic health records," *New England Journal of Medicine*, vol. 363, pp. 501-504, 2010.

[9] A. K. Jha, "Meaningful use of electronic health records: the road ahead," *Jama*, vol. 304, pp. 1709-1710, 2010.

[10] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, p. 1, 2014.

[11] T. B. Murdoch and A. S. Detsky, "The inevitable application of big data to health care," *Jama*, vol. 309, pp. 1351-1352, 2013.

[12] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing big data issues in Scientific Data Infrastructure," in *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, 2013, pp. 48-55.

[13] R. Higdon, W. Haynes, L. Stanberry, E. Stewart, G. Yandl, C. Howard, *et al.*, "Unraveling the Complexities of Life Sciences Data," *Big Data*, vol. 1, pp. 42-50, 2013/03/01 2012.

[14] "The Aging of the U.S. Population and Increased Need for Critical Care Services. Critical Care Workforce Partnership Position Statement," American Association of Critical-Care Nurses, American College of Chest Physicians, American Thoracic Society and Society of Critical Care Medicine. November 2001.

[15] "US Department of Health and Human Services, Agency for Healthcare Research and Quality. National estimates on use of hospitals by children from the HCUP Kids' Inpatient Database (KID). Rockville, MD: Agency for Healthcare Research and Quality," ed, 2009.

[16] S. K. Pasquali, M. L. Jacobs, X. He, S. S. Shah, E. D. Peterson, M. Hall, *et al.*, "Variation in congenital heart surgery costs across hospitals," *Pediatrics*, vol. 133, pp. e553-60, Mar 2014.

[17] L. W. Hayes, E. L. Dobyns, B. DiGiovine, A. M. Brown, S. Jacobson, K. H. Randall, *et al.*, "A multicenter collaborative approach to reducing pediatric codes outside the ICU," *Pediatrics,* vol. 129, pp. e785-91, Mar 2012.
[18] K. L. Brown, D. A. Ridout, A. P. Goldman, A. Hoskote, and D. J. Penny, "Risk factors for long intensive care unit stay after cardiopulmonary bypass in children," *Crit Care Med,* vol. 31, pp. 28-33, Jan 2003.
[19] C. Del Bufalo, A. Morelli, L. Bassein, L. Fasano, C. C. Quarta, A. M. Pacilli, *et al.*, "Severity scores in respiratory intensive care: APACHE II predicted mortality better than SAPS II," *Respiratory care,* vol. 40, pp. 1042-1047, 1995.
[20] L. Fuchs, C. Chronaki, S. Park, V. Novack, Y. Baumfeld, D. Scott, *et al.*, "ICU admission characteristics and mortality rates among elderly and very elderly patients," *Intensive Care Medicine,* vol. 38, pp. 1654-1661, 2012/10/01 2012.
[21] L. Celi, R. Tang, M. Villarroel, G. Davidzon, W. Lester, and H. Chueh, "A Clinical Database-Driven Approach to Decision Support: Predicting Mortality Among Patients with Acute Kidney Injury," *Journal of Healthcare Engineering,* vol. 2, pp. 97-110, 2011.
[22] S. Hunziker, L. A. Celi, J. Lee, and M. D. Howell, "Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients," *Crit Care,* vol. 16, p. R89, May 18 2012.
[23] J. Y. Fagon, A. Novara, F. Stephan, E. Girou, and M. Safar, "Mortality attributable to nosocomial infections in the ICU," *Infect Control Hosp Epidemiol,* vol. 15, pp. 428-34, Jul 1994.
[24] V. J. Ribas, J. C. Lopez, J. C. Ruiz-Rodriguez, A. Ruiz-Sanmartin, J. Rello, and A. Vellido, "On the use of decision trees for ICU outcome prediction in sepsis patients treated with statins," in *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on*, 2011, pp. 37-43.
[25] S. Nemati, A. Malhotra, and G. D. Clifford, "T-wave alternans patterns during sleep in healthy, cardiac disease, and sleep apnea patients," *J Electrocardiol,* vol. 44, pp. 126-30, Mar-Apr 2011.
[26] J.-Y. Fagon, J. Chastre, A. J. Hance, P. Montravers, A. Novara, and C. Gibert, "Nosocomial pneumonia in ventilated patients: A cohort study evaluating attributable mortality and hospital stay," *The American Journal of Medicine,* vol. 94, pp. 281-288, 3// 1993.
[27] S. E. Rooij, A. Govers, J. C. Korevaar, A. Abu-Hanna, M. Levi, and E. Jonge, "Short-term and long-term mortality in very elderly patients admitted to an intensive care unit," *Intensive Care Medicine,* vol. 32, pp. 1039-1044, 2006/07/01 2006.
[28] L. A. Celi, S. Galvin, G. Davidzon, J. Lee, D. Scott, and R. Mark, "A Database-driven Decision Support System: Customized Mortality Prediction," *Journal of Personalized Medicine,* vol. 2, pp. 138-148, 2012.
[29] T. Mandelbaum, D. J. Scott, J. Lee, R. G. Mark, A. Malhotra, S. S. Waikar, *et al.*, "Outcome of critically ill patients with acute kidney injury using the Acute Kidney Injury Network criteria," *Crit Care Med,* vol. 39, pp. 2659-64, Dec 2011.
[30] A. S. Fialho, F. Cismondi, S. M. Vieira, S. R. Reti, J. M. C. Sousa, and S. N. Finkelstein, "Data mining using clinical physiology at discharge to predict ICU

readmissions," *Expert Systems with Applications,* vol. 39, pp. 13158-13165, 12/15/ 2012.
[31] J. Lee and R. G. Mark, "An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care," *Biomed Eng Online,* vol. 9, pp. 9-62, 2010.
[32] C. W. Hug, G. D. Clifford, and A. T. Reisner, "Clinician blood pressure documentation of stable intensive care patients: an intelligent archiving agent has a higher association with future hypotension," *Crit Care Med,* vol. 39, pp. 1006-14, May 2011.
[33] J. Lee, R. Kothari, J. A. Ladapo, D. J. Scott, and L. A. Celi, "Interrogating a clinical database to study treatment of hypotension in the critically ill," *BMJ Open,* vol. 2, 2012.
[34] D. S. Wheeler, M. J. Giaccone, N. Hutchinson, M. Haygood, P. Bondurant, K. Demmel, *et al.*, "A hospital-wide quality-improvement collaborative to reduce catheter-associated bloodstream infections," *Pediatrics,* vol. 128, pp. e995-e1004; quiz e1004-7, Oct 2011.
[35] S. G. Hutchinson, I. Mesters, G. van Breukelen, J. W. Muris, F. J. Feron, S. K. Hammond, *et al.*, "A motivational interviewing intervention to PREvent PAssive Smoke Exposure (PREPASE) in children with a high risk of asthma: design of a randomised controlled trial," *BMC public health,* vol. 13, pp. 1-12, 2013.
[36] S. Kwon, M. Florence, P. Grigas, M. Horton, K. Horvath, M. Johnson, *et al.*, "Creating a learning healthcare system in surgery: Washington State's Surgical Care and Outcomes Assessment Program (SCOAP) at 5 years," *Surgery,* vol. 151, pp. 146-52, Feb 2012.
[37] A. Slater, F. Shann, and G. Pearson, "PIM2: a revised version of the Paediatric Index of Mortality," *Intensive Care Medicine,* vol. 29, pp. 278-285, 2003/02/01 2003.
[38] M. M. Pollack, K. M. Patel, and U. E. Ruttimann, "PRISM III: An updated Pediatric Risk of Mortality score," *Critical Care Medicine,* vol. 24, pp. 743-752, 1996.
[39] F. V. Castello, A. Cassano, P. Gregory, and J. Hammond, "The Pediatric Risk of Mortality (PRISM) Score and Injury Severity Score (ISS) for predicting resource utilization and outcome of intensive care in pediatric trauma," *Crit Care Med,* vol. 27, pp. 985-8, 1999.
[40] M. M. Pollack, U. E. Ruttimann, and P. R. Getson, "Pediatric risk of mortality (PRISM) score," *Crit Care Med,* vol. 16, pp. 1110-6, 1988.
[41] M. M. Pollack, U. E. Ruttimann, and P. R. Getson, "Pediatric risk of mortality (PRISM) score," ed.
[42] L. Anthony Celi, R. G. Mark, D. J. Stone, and R. A. Montgomery, "“Big Data” in the Intensive Care Unit. Closing the Data Loop," *American Journal of Respiratory and Critical Care Medicine,* vol. 187, pp. 1157-1160, 2013.
[43] G. D. Clifford, W. J. Long, G. B. Moody, and P. Szolovits, "Robust parameter extraction for decision support using multimodal intensive care data," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 367, pp. 411-429, January 28, 2009 2009.

[44] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, "Secondary use of EHR: data quality issues and informatics opportunities," *AMIA summits on translational science proceedings,* vol. 2010, p. 1, 2010.
[45] N. G. Weiskopf and C. Weng, "Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research," *Journal of the American Medical Informatics Association,* vol. 20, pp. 144-151, 2013.
[46] S. McPherson, C. Barbosa-Leiker, M. McDonell, D. Howell, and J. Roll, "Longitudinal missing data strategies for substance use clinical trials using generalized estimating equations: an example with a buprenorphine trial," *Human Psychopharmacology: Clinical and Experimental,* vol. 28, pp. 506-515, 2013.
[47] L. R. Zelnick, L. J. Morrison, S. M. Devlin, E. M. Bulger, K. J. Brasel, K. Sheehan, *et al.*, "Addressing the Challenges of Obtaining Functional Outcomes in Traumatic Brain Injury Research: Missing Data Patterns, Timing of Follow-Up, and Three Prognostic Models," *Journal of neurotrauma,* 2014.
[48] A. K. Jha, C. M. DesRoches, E. G. Campbell, K. Donelan, S. R. Rao, T. G. Ferris, *et al.*, "Use of Electronic Health Records in U.S. Hospitals," *New England Journal of Medicine,* vol. 360, pp. 1628-1638, 2009.
[49] G. Molenberghs and E. Lesaffre, "Missing Data," *Encyclopedia of Statistical Sciences,* 2013.
[50] C. Paxton, A. Niculescu-Mizil, and S. Saria, "Developing predictive models using electronic medical records: challenges and pitfalls," *AMIA Annu Symp Proc,* vol. 2013, pp. 1109-15, 2013.
[51] S. I. Goldberg, A. Niemierko, and A. Turchin, "Analysis of data errors in clinical research databases," *AMIA Annu Symp Proc,* pp. 242-6, 2008.
[52] J. A. Evans, "Electronic medical records system," ed: Google Patents, 1999.
[53] S. Fleischer, A. Berg, J. Behrens, O. Kuss, R. Becker, A. Horbach, *et al.*, "Does an additional structured information program during the intensive care unit stay reduce anxiety in ICU patients?: a multicenter randomized controlled trial," *BMC Anesthesiology,* vol. 14, p. 48, 2014.
[54] T. Botsis, G. Hartvigsen, F. Chen, and C. Weng, "Secondary Use of EHR: Data Quality Issues and Informatics Opportunities," *Summit on Translational Bioinformatics,* vol. 2010, pp. 1-5, 03/01 2010.
[55] O. T. Abdala and M. Saeed, "Estimation of missing values in clinical laboratory measurements of ICU patients using a weighted K-nearest neighbors algorithm," in *Computers in Cardiology, 2004*, 2004, pp. 693-696.
[56] S. Hunziker, L. Celi, J. Lee, and M. Howell, "Red cell distribution width improves the simplified acute physiology score for risk prediction in unselected critically ill patients," *Critical Care,* vol. 16, p. R89, 2012.
[57] J. Tian, B. Yu, D. Yu, and S. Ma, "Missing data analyses: a hybrid multiple imputation algorithm using Gray System Theory and entropy based on clustering," *Applied intelligence,* vol. 40, pp. 376-388, 2014.
[58] B. Mitra, M. Fitzgerald, and J. Chan, "The utility of a shock index≥ 1 as an indication for pre-hospital oxygen carrier administration in major trauma," *Injury,* vol. 45, pp. 61-65, 2014.
[59] I. Cho, I. Park, E. Kim, E. Lee, and D. W. Bates, "Using EHR data to predict hospital-acquired pressure ulcers: A prospective study of a Bayesian Network

model," *International journal of medical informatics,* vol. 82, pp. 1059-1067, 2013.
[60] M. M. Pollack, K. M. Patel, and U. E. Ruttimann, "PRISM III: an updated Pediatric Risk of Mortality score," *Crit Care Med,* vol. 24, pp. 743-52, May 1996.
[61] J. Labarère, R. Bertrand, and M. J. Fine, "How to derive and validate clinical prediction models for use in intensive care medicine," *Intensive care medicine,* vol. 40, pp. 513-527, 2014.
[62] D. Macrae, R. Grieve, E. Allen, Z. Sadique, K. Morris, J. Pappachan, *et al.*, "A randomized trial of hyperglycemic control in pediatric intensive care," *New England Journal of Medicine,* vol. 370, pp. 107-118, 2014.
[63] M. Sun, Q. D. Trinh, M. Bianchi, J. Hansen, F. Abdollah, Z. Tian, *et al.*, "Extent of lymphadenectomy does not improve the survival of patients with renal cell carcinoma and nodal metastases: biases associated with the handling of missing data," *BJU international,* vol. 113, pp. 36-42, 2014.
[64] D. Mavridis, A. Chaimani, O. Efthimiou, S. Leucht, and G. Salanti, "Addressing missing outcome data in meta-analysis," *Evidence Based Mental Health,* pp. ebmental-2014-101900, 2014.
[65] K. A. Hallgren and K. Witkiewitz, "Missing data in alcohol clinical trials: a comparison of methods," *Alcoholism: Clinical and Experimental Research,* vol. 37, pp. 2152-2160, 2013.
[66] I. B. Aydilek and A. Arslan, "A hybrid method for imputation of missing values using optimized fuzzy< i> c</i>-means with support vector regression and a genetic algorithm," *Information Sciences,* vol. 233, pp. 25-35, 2013.
[67] B. M. Marlin, D. C. Kale, R. G. Khemani, and R. C. Wetzel, "Unsupervised pattern discovery in electronic health care data using probabilistic clustering models," presented at the Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, New York, NY, USA, 2012.
[68] J. Zhang and S. Gong, "Action categorization with modified hidden conditional random field," *Pattern Recognition,* vol. 43, pp. 197-203, 2010.
[69] H. P. Blumberg, C. Fredericks, F. Wang, J. H. Kalmar, L. Spencer, X. Papademetris, *et al.*, "Preliminary evidence for persistent abnormalities in amygdala volumes in adolescents and young adults with bipolar disorder," *Bipolar Disord,* vol. 7, pp. 570-6, Dec 2005.
[70] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun, "Learning Motion Patterns of People for Compliant Robot Motion," *The International Journal of Robotics Research,* vol. 24, pp. 31-48, January 1, 2005 2005.
[71] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition,* vol. 36, pp. 585-601, 2003.
[72] N. Städler, D. J. Stekhoven, and P. Bühlmann, "Pattern Alternating Maximization Algorithm for Missing Data in High-Dimensional Problems," *Journal of Machine Learning Research,* vol. 15, pp. 1903-1928, 2014.
[73] A. C. Grobler, G. Matthews, and G. Molenberghs, "The impact of missing data on clinical trials: a re-analysis of a placebo controlled trial of (St Johns wort) and sertraline in major depressive disorder," *Psychopharmacology,* vol. 231, pp. 1987-1999, 2014.

[74] P. T. von Hippel, "Should a Normal Imputation Model be Modified to Impute Skewed Variables?," *Sociological Methods & Research,* vol. 42, pp. 105-138, 2013.
[75] A. Mackinnon, "The use and reporting of multiple imputation in medical research - a review," *Journal of Internal Medicine,* vol. 268, pp. 586-593, 2010.
[76] J. M. Jerez, I. Molina, P. J. García-Laencina, E. Alba, N. Ribelles, M. Martín, *et al.*, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial intelligence in medicine,* vol. 50, pp. 105-115, 2010.
[77] P. Jenkins and J. Welton, "Measuring Direct Nursing Cost Per Patient in the Acute Care Setting," *Journal of Nursing Administration,* vol. 44, pp. 257-262, 2014.
[78] L. Fuchs, C. E. Chronaki, S. Park, V. Novack, Y. Baumfeld, D. Scott, *et al.*, "ICU admission characteristics and mortality rates among elderly and very elderly patients," *Intensive Care Med,* vol. 38, pp. 1654-61, Oct 2012.
[79] S. Kim, W. Kim, and R. W. Park, "A Comparison of Intensive Care Unit Mortality Prediction Models through the Use of Data Mining Techniques," *Healthc Inform Res,* vol. 17, pp. 232-243, 12/ 2011.
[80] L. S. S. Wong and J. D. Young, "A comparison of ICU mortality prediction using the APACHE II scoring system and artificial neural networks," *Anaesthesia,* vol. 54, pp. 1048-1054, 1999.
[81] T. Van Nguyen and B. Mishra, "Modeling hospitalization outcomes with random decision trees and bayesian feature selection," *Unpublished, Site: http://www. cs. nyu. edu/mishra/PUBLICATIONS/mypub. html*.
[82] C. Tao, K. Wongsuphasawat, K. Clark, C. Plaisant, B. Shneiderman, and C. G. Chute, "Towards event sequence representation, reasoning and visualization for EHR data," presented at the Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium, Miami, Florida, USA, 2012.
[83] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, "Aligning temporal data by sentinel events: discovering patterns in electronic health records," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2008, pp. 457-466.
[84] H. Syed and A. K. Das, "Identifying Chemotherapy Regimens in Electronic Health Record Data Using Interval-Encoded Sequence Alignment," in *Artificial Intelligence in Medicine*, ed: Springer, 2015, pp. 143-147.
[85] I. J. Casanova, M. Campos, J. M. Juarez, A. Fernandez-Fernandez-Arroyo, and J. A. Lorente, "Using Multivariate Sequential Patterns to Improve Survival Prediction in Intensive Care Burn Unit," in *Artificial Intelligence in Medicine: 15th Conference on Artificial Intelligence in Medicine, AIME 2015, Pavia, Italy, June 17-20, 2015. Proceedings*, H. J. Holmes, R. Bellazzi, L. Sacchi, and N. Peek, Eds., ed Cham: Springer International Publishing, 2015, pp. 277-286.
[86] I. Batal, H. Valizadegan, G. F. Cooper, and M. Hauskrecht, "A pattern mining approach for classifying multivariate temporal data," in *Bioinformatics and Biomedicine (BIBM), 2011 IEEE International Conference on*, 2011, pp. 358-365.

[87] H. Yang and C. C. Yang, "Using Health-Consumer-Contributed Data to Detect Adverse Drug Reactions by Association Mining with Temporal Analysis," *ACM Trans. Intell. Syst. Technol.*, vol. 6, pp. 1-27, 2015.

[88] R. Bellazzi, F. Ferrazzi, and L. Sacchi, "Predictive data mining in clinical medicine: a focus on selected methods and applications," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, pp. 416-430, 2011.

[89] J. L. Warner, A. Zollanvari, Q. Ding, P. Zhang, G. M. Snyder, and G. Alterovitz, "Temporal phenome analysis of a large electronic health record cohort enables identification of hospital-acquired complications," *Journal of the American Medical Informatics Association*, vol. 20, pp. e281-e287, 2013.

[90] T. H. McCoy, V. M. Castro, A. Cagan, A. M. Roberson, I. S. Kohane, and R. H. Perlis, "Sentiment Measured in Hospital Discharge Notes Is Associated with Readmission and Mortality Risk: An Electronic Health Record Study," *PloS one*, vol. 10, p. e0136341, 2015.

[91] X. Cai, O. Perez-Concha, E. Coiera, F. Martin-Sanchez, R. Day, D. Roffe, *et al.*, "Real-time prediction of mortality, readmission, and length of stay using electronic health record data," *Journal of the American Medical Informatics Association*, p. ocv110, 2015.

[92] D. Toddenroth, T. Ganslandt, I. Castellanos, H.-U. Prokosch, and T. Bürkle, "Employing heat maps to mine associations in structured routine care data," *Artificial intelligence in medicine*, 2013.

[93] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Medical Concept Representation Learning from Electronic Health Records and its Application on Heart Failure Prediction," *arXiv preprint arXiv:1602.03686*, 2016.

[94] A. Singh, G. Nadkarni, O. Gottesman, S. B. Ellis, E. P. Bottinger, and J. V. Guttag, "Incorporating temporal EHR data in predictive models for risk stratification of renal function deterioration," *Journal of biomedical informatics*, vol. 53, pp. 220-228, 2015.

[95] F. Doshi-Velez, Y. Ge, and I. Kohane, "Comorbidity clusters in autism spectrum disorders: an electronic health record time-series analysis," *Pediatrics*, vol. 133, pp. e54-e63, 2014.

[96] Z. Liu, L. Wu, and M. Hauskrecht, "Modeling clinical time series using Gaussian process sequences," in *SIAM international conference on data mining*, 2013, pp. 623-631.

[97] G. Stiglic, A. Davey, and Z. Obradovic, "Temporal Evaluation of Risk Factors for Acute Myocardial Infarction Readmissions," in *Healthcare Informatics (ICHI), 2013 IEEE International Conference on*, 2013, pp. 557-562.

[98] H.-c. Lin, V. Baracos, R. Greiner, and J. Y. Chun-nam, "Learning patient-specific cancer survival distributions as a sequence of dependent regressors," in *Advances in Neural Information Processing Systems*, 2011, pp. 1845-1853.

[99] R. V. Andreão and J. Boudy, "A comparison of wavelet transforms through an HMM based ECG segmentation and classification system," in *The Proceeding of IASTED Conference on Biomedical Engineering Innsbruck, Austria February*, 2006.

[100] R. V. Andreao, B. Dorizzi, and J. Boudy, "ECG signal analysis through hidden Markov models," *Biomedical Engineering, IEEE Transactions on,* vol. 53, pp. 1541-1549, 2006.
[101] G. de Lannoy, D. François, J. Delbeke, and M. Verleysen, "Weighted conditional random fields for supervised interpatient heartbeat classification," *Biomedical Engineering, IEEE Transactions on,* vol. 59, pp. 241-247, 2012.
[102] H. Manabe and Z. Zhang, "Multi-stream HMM for EMG-based speech recognition," in *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE,* 2004, pp. 4389-4392.
[103] J. Thomas, C. Rose, and F. Charpillet, "A multi-HMM approach to ECG segmentation," in *Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on*, 2006, pp. 609-616.
[104] W. D. Penny and S. J. Roberts, "Dynamic Models for Nonstationary Signal Segmentation," *Computers and Biomedical Research,* vol. 32, pp. 483-502, 12// 1999.
[105] D. Husmeier, R. Dybowski, and S. Roberts, "Probabilistic modeling in bioinformatics and medical informatics," 2005.
[106] P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," *Nature Reviews Genetics,* vol. 13, pp. 395-405, 2012.
[107] I. Batal and C. San Ramon, "Temporal data mining for healthcare data," *Healthcare Data analytics. Boca Raton, FL: Chapman and Hall/CRC,* pp. 379-402, 2015.
[108] G. D. Clifford, W. J. Long, G. B. Moody, and P. Szolovits, "Robust parameter extraction for decision support using multimodal intensive care data," *Philos Trans A Math Phys Eng Sci,* vol. 367, pp. 411-29, Jan 28 2009.
[109] R. Martis, C. Chakraborty, and A. Ray, "Wavelet-based Machine Learning Techniques for ECG Signal Analysis," in *Machine Learning in Healthcare Informatics*. vol. 56, S. Dua, U. R. Acharya, and P. Dua, Eds., ed: Springer Berlin Heidelberg, 2014, pp. 25-45.
[110] A. Faiola and C. Newlon, "Advancing critical care in the ICU: a human-centered biomedical data visualization systems," presented at the Proceedings of the 2011th international conference on Ergonomics and health aspects of work with computers, Orlando, FL, USA, 2011.
[111] T. A. Lasko, J. C. Denny, and M. A. Levy, "Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data," *PloS one,* vol. 8, p. e66341, 2013.
[112] R. Miotto, L. Li, and J. T. Dudley, "Deep Learning to Predict Patient Future Diseases from the Electronic Health Records," in *European Conference on Information Retrieval*, 2016, pp. 768-774.
[113] B. K. Beaulieu-Jones and C. S. Greene, "Semi-Supervised Learning of the Electronic Health Record with Denoising Autoencoders for Phenotype Stratification," *bioRxiv,* p. 039800, 2016.
[114] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," *Scientific reports,* vol. 6, 2016.

[115] D. Windridge and M. Bober, "A kernel-based framework for medical big-data analytics," in *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics*, ed: Springer, 2014, pp. 197-208.
[116] G. S. Babu, P. Zhao, and X.-L. Li, "Deep Convolutional Neural Network Based Regression Approach for Estimation of Remaining Useful Life," in *International Conference on Database Systems for Advanced Applications*, 2016, pp. 214-228.
[117] T. Tran, T. D. Nguyen, D. Phung, and S. Venkatesh, "Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM)," *Journal of biomedical informatics*, vol. 54, pp. 96-105, 2015.
[118] D. C. Mocanu, H. B. Ammar, D. Lowet, K. Driessens, A. Liotta, G. Weiss, *et al.*, "Factored four way conditional restricted boltzmann machines for activity recognition," *Pattern Recognition Letters*, vol. 66, pp. 100-108, 2015.
[119] P. Cao, X. Liu, H. Bao, J. Yang, and D. Zhao, "Restricted Boltzmann machines based oversampling and semi-supervised learning for false positive reduction in breast CAD," *Bio-Medical Materials and Engineering*, vol. 26, pp. S1541-S1547, 2015.
[120] E. Choi, M. T. Bahadori, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," *arXiv preprint arXiv:1511.05942*, 2015.
[121] E. Choi, A. Schuetz, W. F. Stewart, and J. Sun, "Using recurrent neural network models for early detection of heart failure onset," *Journal of the American Medical Informatics Association*, p. ocw112, 2016.
[122] A. Jagannatha and H. Yu, "Bidirectional Recurrent Neural Networks for Medical Event Detection in Electronic Health Records," *arXiv preprint arXiv:1606.07953*, 2016.
[123] Z. C. Lipton, D. C. Kale, C. Elkan, and R. Wetzell, "Learning to Diagnose with LSTM Recurrent Neural Networks," *arXiv preprint arXiv:1511.03677*, 2015.
[124] A. R. Johansen, J. Jin, T. Maszczyk, J. Dauwels, S. S. Cash, and M. B. Westover, "Epileptiform spike detection via convolutional neural networks," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 754-758.
[125] C. Sideris, H. Kalantarian, E. Nemati, and M. Sarrafzadeh, "Building Continuous Arterial Blood Pressure Prediction Models Using Recurrent Networks," in *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2016, pp. 1-5.
[126] S. Stober, A. Sternin, A. M. Owen, and J. A. Grahn, "Deep Feature Learning for EEG Recordings," *arXiv preprint arXiv:1511.04306*, 2015.
[127] S. Stober, D. J. Cameron, and J. A. Grahn, "Using Convolutional Neural Networks to Recognize Rhythm Stimuli from Electroencephalography Recordings," in *Advances in Neural Information Processing Systems*, 2014, pp. 1449-1457.
[128] G. Hripcsak and D. J. Albers, "Next-generation phenotyping of electronic health records," *Journal of the American Medical Informatics Association*, vol. 20, pp. 117-121, 2013.

[129] N. E. Adler and W. W. Stead, "Patients in context—EHR capture of social and behavioral determinants of health," *New England Journal of Medicine,* vol. 372, pp. 698-701, 2015.
[130] J. Andreu-Perez, C. C. Poon, R. D. Merrifield, S. T. Wong, and G.-Z. Yang, "Big data for health," *IEEE journal of biomedical and health informatics,* vol. 19, pp. 1193-1208, 2015.
[131] O. Gottesman, H. Kuivaniemi, G. Tromp, W. A. Faucett, R. Li, T. A. Manolio, *et al.*, "The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future," *Genet Med,* vol. 15, pp. 761-71, Oct 2013.
[132] B. S. Glicksberg, L. Li, W.-Y. Cheng, K. Shameer, J. Hakenberg, R. Castellanos, *et al.*, "An integrative pipeline for multi-modal discovery of disease relationships," in *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 2015, p. 407.
[133] R. Akerkar, "Towards an Intelligent Integrated Approach for Clinical Decision Support," in *Managing Big Data Integration in the Public Sector*, ed: IGI Global, 2016, pp. 187-205.
[134] R. R. Andridge and R. J. A. Little, "A Review of Hot Deck Imputation for Survey Non-response," *International statistical review = Revue internationale de statistique,* vol. 78, pp. 40-64, 2010.
[135] R. J. Little, "Modeling the drop-out mechanism in repeated-measures studies," *Journal of the American Statistical Association,* vol. 90, pp. 1112-1121, 1995.
[136] R. J. Little, "A test of missing completely at random for multivariate data with missing values," *Journal of the American Statistical Association,* vol. 83, pp. 1198-1202, 1988.
[137] S. Nakagawa and R. P. Freckleton, "Missing inaction: the dangers of ignoring missing data," *Trends in Ecology & Evolution,* vol. 23, pp. 592-596, 11// 2008.
[138] F. Cismondi, A. S. Fialho, S. M. Vieira, S. R. Reti, J. M. Sousa, and S. N. Finkelstein, "Missing data in medical databases: impute, delete or classify?," *Artif Intell Med,* vol. 58, pp. 63-72, May 2013.
[139] H. A. Kiers and J. M. ten Berge, "Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices in two or more populations," *Psychometrika,* vol. 54, pp. 467-473, 1989.
[140] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on neural networks,* vol. 16, pp. 645-678, 2005.
[141] B. Michiels, G. Molenberghs, L. Bijnens, T. Vangeneugden, and H. Thijs, "Selection models and pattern-mixture models to analyse longitudinal quality of life data subject to drop-out," *Statistics in Medicine,* vol. 21, pp. 1023-1041, 2002.
[142] M. O'Kelly and B. Ratitch, "Analyses under Missing-not-at-random Assumptions," *Clinical Trials with Missing Data: A Guide for Practitioners,* pp. 257-368, 2014.
[143] C. Romano, "Applying copula function to risk management," in *Capitalia, Italy. http://www. icer. it/workshop/Romano. pdf*, 2002.
[144] J. Gatz, "Master Theses: Properties and Applications of the Student T Copula," 2007.
[145] C. Smart and C. Director, "Beyond Correlation: Don’t Use the Formula that Killed Wall Street."

[146] E. Bouyé, V. Durrleman, A. Nikeghbali, G. Riboulet, and T. Roncalli, "Copulas for finance-a reading guide and some applications," 2000.
[147] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC bioinformatics,* vol. 7, p. 91, 2006.
[148] L. Zheng, D. Watson, B. Johnston, R. L. Clark, R. Edrada-Ebel, and W. Elseheri, "A chemometric study of chromatograms of tea extracts by correlation optimization warping in conjunction with PCA, support vector machines and random forest data modeling," *Analytica Chimica Acta,* vol. 642, pp. 257-265, 2009.
[149] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, *et al.*, "Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): a public-access intensive care unit database," *Critical care medicine,* vol. 39, p. 952, 2011.
[150] A. E. Johnson, T. J. Pollard, L. Shen, H. L. Li-wei, M. Feng, M. Ghassemi, *et al.*, "MIMIC-III, a freely accessible critical care database," *Scientific data,* vol. 3, p. 160035, 2016.
[151] J. H. Steiger, "Tests for comparing elements of a correlation matrix," *Psychological Bulletin,* vol. 87, p. 245, 1980.
[152] V. K. Moitra, C. Guerra, W. T. Linde-Zwirble, and H. Wunsch, "Relationship between ICU length of stay and long-term mortality for elderly ICU survivors," *Critical care medicine,* vol. 44, p. 655, 2016.
[153] W. A. Knaus, D. P. Wagner, J. E. Zimmerman, and E. A. Draper, "Variations in mortality and length of stay in intensive care units," *Annals of Internal Medicine,* vol. 118, pp. 753-761, 1993.
[154] T. Williams, K. Ho, G. Dobb, J. Finn, M. Knuiman, and S. Webb, "Effect of length of stay in intensive care unit on hospital and long-term mortality of critically ill adult patients," *British journal of anaesthesia,* vol. 104, pp. 459-464, 2010.
[155] É. Azoulay, E. Canet, E. Raffoux, E. Lengliné, V. Lemiale, F. Vincent, *et al.*, "Dexamethasone in patients with acute lung injury from acute monocytic leukemia," *European Respiratory Journal,* pp. erj00577-2011, 2011.
[156] F. L. Ferreira, D. P. Bota, A. Bross, C. Mélot, and J.-L. Vincent, "Serial evaluation of the SOFA score to predict outcome in critically ill patients," *Jama,* vol. 286, pp. 1754-1758, 2001.
[157] L. Mayaud, P. S. Lai, G. D. Clifford, L. Tarassenko, L. A. G. Celi, and D. Annane, "Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension," *Critical care medicine,* vol. 41, p. 954, 2013.
[158] M. L. Vold, U. Aasebø, T. Wilsgaard, and H. Melbye, "Low oxygen saturation and mortality in an adult cohort: the Tromsø study," *BMC pulmonary medicine,* vol. 15, p. 9, 2015.
[159] Y. Chen and H. Yang, "Heterogeneous postsurgical data analytics for predictive modeling of mortality risks in intensive care units," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th annual international conference of the IEEE*, 2014, pp. 4310-4314.

[160] R. Pirracchio, "Mortality Prediction in the ICU Based on MIMIC-II Results from the Super ICU Learner Algorithm (SICULA) Project," in *Secondary Analysis of Electronic Health Records*, ed: Springer, 2016, pp. 295-313.
[161] Á. Castellanos-Ortega, B. Suberviola, L. A. García-Astudillo, M. S. Holanda, F. Ortiz, J. Llorca, *et al.*, "Impact of the Surviving Sepsis Campaign protocols on hospital length of stay and mortality in septic shock patients: results of a three-year follow-up quasi-experimental study," *Critical care medicine,* vol. 38, pp. 1036-1043, 2010.
[162] J.-L. Vincent, R. Moreno, J. Takala, S. Willatts, A. De Mendonça, H. Bruining, *et al.*, "The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure," ed: Springer, 1996.
[163] J. C. Schefold, C. Storm, S. Bercker, R. Pschowski, M. Oppert, A. Krüger, *et al.*, "Inferior vena cava diameter correlates with invasive hemodynamic measures in mechanically ventilated intensive care unit patients with sepsis," *The Journal of emergency medicine,* vol. 38, pp. 632-637, 2010.
[164] J. S. Calvert, D. A. Price, U. K. Chettipally, C. W. Barton, M. D. Feldman, J. L. Hoffman, *et al.*, "A computational approach to early sepsis detection," *Computers in biology and medicine,* vol. 74, pp. 69-73, 2016.
[165] T. Desautels, J. Calvert, J. Hoffman, Q. Mao, M. Jay, G. Fletcher, *et al.*, "Using Transfer Learning for Improved Mortality Prediction in a Data-Scarce Hospital Setting," *Biomedical informatics insights,* vol. 9, p. 1178222617712994, 2017.
[166] S. Purushotham, C. Meng, Z. Che, and Y. Liu, "Benchmark of Deep Learning Models on Large Healthcare MIMIC Datasets," *arXiv preprint arXiv:1710.08531,* 2017.
[167] B. Shickel, T. J. Loftus, L. Adhikari, T. Ozrazgat-Baslanti, A. Bihorac, and P. Rashidi, "DeepSOFA: A Continuous Acuity Score for Critically Ill Patients using Clinically Interpretable Deep Learning," 2018.
[168] T. Desautels, J. Calvert, J. Hoffman, M. Jay, Y. Kerem, L. Shieh, *et al.*, "Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach," *JMIR medical informatics,* vol. 4, 2016.
[169] L. A. García-Escudero, A. Gordaliza, C. Matrán, and A. Mayo-Iscar, "A review of robust clustering methods," *Advances in Data Analysis and Classification,* vol. 4, pp. 89-109, 2010// 2010.
[170] J. Lee and R. Mark, "An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care," *BioMedical Engineering OnLine,* vol. 9, pp. 1-17, 2010/10/25 2010.
[171] J. D. Lafferty, A. McCallum, and F. C. N. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," presented at the Proceedings of the Eighteenth International Conference on Machine Learning, 2001.
[172] W. Yang, L. Kia-Fock, and W. Jian-Kang, "A dynamic conditional random field model for foreground and shadow segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 28, pp. 279-289, 2006.
[173] C.-H. Lee, M. Schmidt, A. Murtha, A. Bistritz, J. Sander, and R. Greiner, "Segmenting Brain Tumors with Conditional Random Fields and Support Vector

Machines," in *Computer Vision for Biomedical Image Applications*. vol. 3765, Y. Liu, T. Jiang, and C. Zhang, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 469-478.
[174] K. Held, E. R. Kops, B. J. Krause, W. M. I. I. I. Wells, R. Kikinis, and H. W. Muller-Gartner, "Markov random field segmentation of brain MR images," *Medical Imaging, IEEE Transactions on*, vol. 16, pp. 878-886, 1997.
[175] G. Luo and W. Min, "Subject-adaptive real-time sleep stage classification based on conditional random field," *AMIA Annu Symp Proc*, pp. 488-92, 2007.
[176] B. Wang, X. Wang, J. Zou, F. Kawana, and M. Nakamura, "Automatic determination of sleep stage through bio-neurological signals contaminated with artifacts by a conditional probability of the knowledge base," *Artificial Life and Robotics*, vol. 12, pp. 270-275, 2008/03/01 2008.
[177] I. C. Chang and C. L. Huang, "Skeleton-basedWalking Motion Analysis Using Hidden Markov Model and Active Shape Models," *Journal of Information Science and Engineering*, vol. 17, pp. 371-403, 2001.
[178] T. Wenlong and E. S. Sazonov, "Highly accurate classification of postures and activities by a shoe-based monitor through classification with rejection," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, 2012, pp. 2611-2614.
[179] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, pp. 1-43, 2011.
[180] J. Venugopalan, N. Chanani, K. Maher, and M. D. Wang, "Novel Data Imputation for Multiple Types of Missing Data in Intensive Care Units," *Journal of Biomedical and Health Informatics*, 2017 (Accepted).
[181] J. Lafferty, A. McCallum, and F. C. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," 2001.
[182] J. Lafferty, X. Zhu, and Y. Liu, "Kernel conditional random fields: representation and clique selection," in *Proceedings of the twenty-first international conference on Machine learning*, 2004, p. 64.
[183] P. C. J. M. Hammersley, "Markov field on finite graphs and lattices (1971) ".
[184] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *J Bioinform Comput Biol*, vol. 3, pp. 185-205, Apr 2005.
[185] A. L. Rosenberg and C. Watts, "Patients readmitted to ICUs: a systematic review of risk factors and outcomes," *Chest*, vol. 118, pp. 492-502, 2000.
[186] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, pp. 1721-1730.
[187] L. M. Chen, C. M. Martin, S. P. Keenan, and W. J. Sibbald, "Patients readmitted to the intensive care unit during the same hospitalization: clinical features and outcomes," *Critical care medicine*, vol. 26, pp. 1834-1841, 1998.
[188] S. Farasat, J. D. Possick, and C. L. Rochester, "Clinical Features And Discharge Characteristics Of Patients Readmitted Within 30 Days Following Index Admission For COPD Exacerbation At Yale-New Haven Hospital," in *A41. THE SPECTRUM COPD CARE: FROM IDENTIFICATION TO POLICY*, ed: American Thoracic Society, 2016, pp. A1526-A1526.

[189] Y.-W. Lin, Y. Zhou, F. Faghri, M. J. Shaw, and R. H. Campbell, "Analysis and Prediction of Unplanned Intensive Care Unit Readmission using Recurrent Neural Networks with Long Short-Term Memory," *bioRxiv,* p. 385518, 2018.
[190] Y. Xue, D. Klabjan, and Y. Luo, "Predicting ICU readmission using grouped physiological and medication trends," *Artificial intelligence in medicine,* 2018.
[191] N. A. Halpern and S. M. Pastores, "Critical care medicine in the United States 2000–2005: An analysis of bed numbers, occupancy rates, payer mix, and costs*," *Critical care medicine,* vol. 38, pp. 65-71, 2010.
[192] D. C. Angus, W. T. Linde-Zwirble, C. A. Sirio, A. J. Rotondi, L. Chelluri, R. C. Newbold, *et al.*, "The effect of managed care on ICU length of stay: implications for Medicare," *Jama,* vol. 276, pp. 1075-1082, 1996.
[193] A. W. Wu, P. Pronovost, and L. Morlock, "ICU incident reporting systems," *Journal of critical care,* vol. 17, pp. 86-94, 2002.
[194] M. Young and J. Birkmeyer, "Potential reduction in mortality rates using an intensivist model to manage intensive care units," *Effective clinical practice: ECP,* vol. 3, pp. 284-289, 1999.
[195] J. Rapoport, D. Teres, S. Lemeshow, J. S. Avrunin, and R. Haber, "Explaining variability of cost using a severity-of-illness measure for ICU patients," *Medical care,* vol. 28, pp. 338-348, 1990.
[196] J. Rapoport, D. Teres, S. Lemeshow, and S. Gehlbach, "A method for assessing the clinical performance and cost-effectiveness of intensive care units: a multicenter inception cohort study," *Critical care medicine,* vol. 22, pp. 1385-1391, 1994.
[197] J. Venugopalan, Z. Zhang, N. Chanani, K. Maher, and M. D. Wang, "Time-Series Data Analysis to Predict Adverse Events in the Intensive Care Unit " *Unpublished,* 2017.
[198] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 27, pp. 1226-1238, 2005.
[199] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of the eighteenth international conference on machine learning, ICML*, 2001, pp. 282-289.
[200] D. A. Gruenberg, W. Shelton, S. L. Rose, A. E. Rutter, S. Socaris, and G. McGee, "Factors Influencing Length of Stay in the Intensive Care Unit," *American Journal of Critical Care,* vol. 15, pp. 502-509, September 1, 2006 2006.
[201] R. Banerjee, J. M. Naessens, E. G. Seferian, O. Gajic, J. P. Moriarty, M. G. Johnson, *et al.*, "Economic implications of nighttime attending intensivist coverage in a medical intensive care unit," *Crit Care Med,* vol. 39, pp. 1257-62, Jun 2011.
[202] A. Donati, S. Loggi, J.-C. Preiser, G. Orsetti, C. Munch, V. Gabbanelli, *et al.*, "Goal-directed intraoperative therapy reduces morbidity and length of hospital stay in high-risk surgical patients," *CHEST Journal,* vol. 132, pp. 1817-1824, 2007.

[203] J. E. Zimmerman, A. A. Kramer, D. S. McNair, and F. M. Malila, "Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients," *Crit Care Med,* vol. 34, pp. 1297-310, May 2006.
[204] J.-L. Vincent and R. Moreno, "Clinical review: Scoring systems in the critically ill," *Critical Care,* vol. 14, p. 207, 2010.
[205] L. Turgeman, J. H. May, and R. Sciulli, "Insights from a machine learning model for predicting the hospital Length of Stay (LOS) at the time of admission," *Expert Systems with Applications,* vol. 78, pp. 376-385, 7/15/ 2017.
[206] C. Ngufor, D. Murphree, S. Upadhyaya, N. Madde, J. Pathak, R. Carter, *et al.*, "Predicting Prolonged Stay in the ICU Attributable to Bleeding in Patients Offered Plasma Transfusion," *AMIA Annual Symposium Proceedings,* vol. 2016, pp. 954-963, 02/10 2016.
[207] M. T. Chuang, Y. h. Hu, and C. L. Lo, "Predicting the prolonged length of stay of general surgery patients: a supervised learning approach," *International Transactions in Operational Research,* 2016.
[208] R. Houthooft, J. Ruyssinck, J. van der Herten, S. Stijven, I. Couckuyt, B. Gadeyne, *et al.*, "Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores," *Artificial Intelligence in Medicine,* vol. 63, pp. 191-207, 3// 2015.
[209] M. Rowan, T. Ryan, F. Hegarty, and N. O'Hare, "The use of artificial neural networks to stratify the length of stay of cardiac patients based on preoperative and initial postoperative factors," *Artificial Intelligence in Medicine,* vol. 40, pp. 211-221, 7// 2007.
[210] J. E. Zimmerman, A. A. Kramer, D. S. McNair, F. M. Malila, and V. L. Shaffer, "Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV," *Crit Care Med,* vol. 34, pp. 2517-29, Oct 2006.
[211] M. Verduijn, N. Peek, F. Voorbraak, E. De Jonge, and B. de Mol, "Dichotomization of ICU length of stay based on model calibration," in *Conference on Artificial Intelligence in Medicine in Europe*, 2005, pp. 67-76.
[212] A. Seth Kapadia, W. Chan, R. Sachdeva, L. A. Moye, and L. S. Jefferson, "Predicting duration of stay in a pediatric intensive care unit: A Markovian approach," *European Journal of Operational Research,* vol. 124, pp. 353-359, 7/16/ 2000.
[213] T. A. Williams, K. M. Ho, G. J. Dobb, J. C. Finn, M. Knuiman, and S. A. R. Webb, "Effect of length of stay in intensive care unit on hospital and long-term mortality of critically ill adult patients," *BJA: British Journal of Anaesthesia,* vol. 104, pp. 459-464, 2010.
[214] Z.-H. Zhou, *Ensemble methods: foundations and algorithms*: CRC press, 2012.
[215] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple classifier systems*, ed: Springer, 2000, pp. 1-15.
[216] I. Batal, D. Fradkin, J. Harrison, F. Moerchen, and M. Hauskrecht, "Mining recent temporal patterns for event detection in multivariate time series data," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 280-288.

[217] L. Zhou and G. Hripcsak, "Temporal reasoning with medical data—a review with emphasis on medical natural language processing," *Journal of biomedical informatics*, vol. 40, pp. 183-202, 2007.
[218] J. Venugopalan, N. Chanani, K. Maher, and D. W. May, "Combination of Static and Temporal Data Analysis to Predict Mortality and Readmission in the Intensive Care," presented at the Engineering in Medicine and Biology Society (EMBC), 2017 Annual International Conference of the IEEE, 2017 (Submitted). Korea, 2017.
[219] H. Li, X. Li, M. Ramanathan, and A. Zhang, "Identifying informative risk factors and predicting bone disease progression via deep belief networks," *Methods*, vol. 69, pp. 257-265, 2014.
[220] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, "Recurrent neural networks for multivariate time series with missing values," *arXiv preprint arXiv:1606.01865*, 2016.
[221] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, pp. 1735-1780, 1997.
[222] T.-H. Wen, M. Gasic, N. Mrksic, P.-H. Su, D. Vandyke, and S. Young, "Semantically conditioned lstm-based natural language generation for spoken dialogue systems," *arXiv preprint arXiv:1508.01745*, 2015.
[223] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, "Dropout improves recurrent neural networks for handwriting recognition," in *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, 2014, pp. 285-290.
[224] A. Graves, N. Jaitly, and A.-r. Mohamed, "Hybrid speech recognition with deep bidirectional LSTM," in *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, 2013, pp. 273-278.
[225] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning," *Nat Biotechnol*, vol. 33, pp. 831-8, Aug 2015.
[226] H. R. Hassanzadeh and M. D. Wang, "DeeperBind: Enhancing Prediction of Sequence Specificities of DNA Binding Proteins," *arXiv preprint arXiv:1611.05777*, 2016.
[227] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *bioinformatics*, vol. 23, pp. 2507-2517, 2007.
[228] G. Doquire and M. Verleysen, "Feature selection with missing data using mutual information estimators," *Neurocomputing*, vol. 90, pp. 3-11, 2012.
[229] Q. Lou and Z. Obradovic, "Margin-Based Feature Selection in Incomplete Data," in *AAAI*, 2012, pp. 1040-1046.
[230] H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Interspeech*, 2014, pp. 338-342.
[231] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *arXiv preprint arXiv:1402.1128*, 2014.
[232] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[233] N. Japkowicz and M. Shah, *Evaluating learning algorithms: a classification perspective*: Cambridge University Press, 2011.
[234] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," 2011.
[235] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, "Distilling knowledge from deep networks with applications to healthcare domain," *arXiv preprint arXiv:1512.03542*, 2015.
[236] V. K. Moitra, C. Guerra, W. T. Linde-Zwirble, and H. Wunsch, "Relationship Between ICU Length of Stay and Long-Term Mortality for Elderly ICU Survivors," *Critical care medicine*, vol. 44, pp. 655-662, 2016.
[237] A. K Szilágyi, C. Diószeghy, G. Fritúz, J. Gál, and K. Varga, "Shortening the length of stay and mechanical ventilation time by using positive suggestions via MP3 players for ventilated patients," *Interventional medicine & applied science*, vol. 6, pp. 3-15, 2014.
[238] M. M. Stecker, M. Stecker, and J. Falotico, "Predictive model of length of stay and discharge destination in neuroscience admissions," *Surgical neurology international*, vol. 8, pp. 17-17, 2017.
[239] R. Tiruvoipati, J. Botha, J. Fletcher, H. Gangopadhyay, M. Majumdar, S. Vij, *et al.*, "Intensive care discharge delay is associated with increased hospital length of stay: A multicentre prospective observational study," *PLOS ONE*, vol. 12, p. e0181827, 2017.
[240] Y. Arabi, S. Venkatesh, S. Haddad, A. A. Shimemeri, and S. A. Malik, "A prospective study of prolonged stay in the intensive care unit: predictors and impact on resource utilization," *International Journal for Quality in Health Care*, vol. 14, pp. 403-410, 2002.
[241] A. H. Choudhuri, M. Chakravarty, and R. Uppal, "Influence of Admission Source on the Outcome of Patients in an Intensive Care Unit," *Indian journal of critical care medicine : peer-reviewed, official publication of Indian Society of Critical Care Medicine*, vol. 21, pp. 213-217, 2017.
[242] K. N. Brown, J. P. Leigh, H. Kamran, S. M. Bagshaw, R. A. Fowler, P. M. Dodek, *et al.*, "Transfers from intensive care unit to hospital ward: a multicentre textual analysis of physician progress notes," *Critical Care*, vol. 22, p. 19, January 28 2018.
[243] D. E. Corl, T. S. Yin, M. E. Mills, T. L. Spencer, L. Greenfield, E. Beauchemin, *et al.*, "Evaluation of point-of-care blood glucose measurements in patients with diabetic ketoacidosis or hyperglycemic hyperosmolar syndrome admitted to a critical care unit," *Journal of diabetes science and technology*, vol. 7, pp. 1265-1274, 2013.
[244] E. Kipnis, D. Ramsingh, M. Bhargava, E. Dincer, M. Cannesson, A. Broccard, *et al.*, "Monitoring in the Intensive Care," *Critical Care Research and Practice*, vol. 2012, p. 20, 2012.
[245] A. B. Böhmer, K. S. Just, R. Lefering, T. Paffrath, B. Bouillon, R. Joppich, *et al.*, "Factors influencing lengths of stay in the intensive care unit for surviving trauma patients: a retrospective analysis of 30,157 cases," *Critical care (London, England)*, vol. 18, pp. R143-R143, 2014.

[246] P. Rajendram, S. Kamat, V. Park, N. Kostelecky, L. Voigt, S. Chawla, *et al.*, "423: ACQUIRED METHEMOGLOBINEMIA IN CANCER PATIENTS ADMITTED TO THE ICU CAUSES AND OUTCOMES," *Critical Care Medicine,* vol. 40, pp. 1-328, 2012.
[247] Y. Sakr, S. Lobo, S. Knuepfer, E. Esser, M. Bauer, U. Settmacher, *et al.*, "Anemia and blood transfusion in a surgical intensive care unit," *Critical care (London, England),* vol. 14, pp. R92-R92, 2010.
[248] Y. Y. Ding, B. Kader, C. L. Christiansen, and D. R. Berlowitz, "Hemoglobin Level and Hospital Mortality Among ICU Patients With Cardiac Disease Who Received Transfusions," *Journal of the American College of Cardiology,* vol. 66, pp. 2510-2518, 2015/12/08/ 2015.
[249] C. Chelazzi, E. Pettini, G. Villa, and A. R. De Gaudio, "Epidemiology, associated factors and outcomes of ICU-acquired infections caused by Gram-negative bacteria in critically ill patients: an observational, retrospective study," *BMC anesthesiology,* vol. 15, pp. 125-125, 2015.
[250] J. Hopman, A. Tostmann, H. Wertheim, M. Bos, E. Kolwijck, R. Akkermans, *et al.*, "Reduced rate of intensive care unit acquired gram-negative bacilli after removal of sinks and introduction of 'water-free' patient care," *Antimicrobial resistance and infection control,* vol. 6, pp. 59-59, 2017.
[251] J. Lee, E. de Louw, M. Niemi, R. Nelson, R. G. Mark, L. A. Celi, *et al.*, "Association between fluid balance and survival in critically ill patients," *Journal of internal medicine,* vol. 277, pp. 468-477, 2015.
[252] C. Cordemans, I. De Laet, N. Van Regenmortel, K. Schoonheydt, H. Dits, W. Huber, *et al.*, "Fluid management in critically ill patients: the role of extravascular lung water, abdominal hypertension, capillary leak, and fluid balance," *Annals of intensive care,* vol. 2, pp. S1-S1, 2012.
[253] D. Armstrong-Briley, N. S. T. Hozhabri, K. Armstrong, J. Puthottile, R. Benavides, and S. Beal, "Comparison of length of stay and outcomes of patients with positive versus negative blood culture results," *Proceedings (Baylor University. Medical Center),* vol. 28, pp. 10-13, 2015.
[254] H. Romo, A. Amaral, and J. Vincent, "Effect of patient sex on intensive care unit survival," *Archives of Internal Medicine,* vol. 164, pp. 61-65, 2004.
[255] K. Tobi and F. Amadasun, "Prolonged stay in the Intensive Care Unit of a tertiary hospital in Nigeria: Predisposing factors and outcome," *African Journal of Medical and Health Sciences,* vol. 14, pp. 56-60, January 1, 2015 2015.
[256] J.-C. Preiser, J. G. Chase, R. Hovorka, J. I. Joseph, J. S. Krinsley, C. De Block, *et al.*, "Glucose Control in the ICU: A Continuing Story," *Journal of diabetes science and technology,* vol. 10, pp. 1372-1381, 2016.
[257] R. Rajendran and G. Rayman, "Point-of-care blood glucose testing for diabetes care in hospitalized patients: an evidence-based review," *Journal of diabetes science and technology,* vol. 8, pp. 1081-1090, 2014.
[258] L. J. Engele, M. Straat, I. H. M. van Rooijen, K. M. K. de Vooght, O. L. Cremer, M. J. Schultz, *et al.*, "Transfusion of platelets, but not of red blood cells, is independently associated with nosocomial infections in the critically ill," *Annals of intensive care,* vol. 6, pp. 67-67, 2016.

[259] M. T. Beigmohammadi, Z. Hussain Khan, S. Samadi, A. Mahmoodpoor, A. Fotouhi, A. Rahimiforoushani, *et al.*, "Role of Hematocrit Concentration on Successful Extubation in Critically Ill Patients in the Intensive Care Units," *Anesthesiology and pain medicine,* vol. 6, pp. e32904-e32904, 2016.
[260] J. S. You, Y. S. Park, H. S. Chung, H. S. Lee, Y. Joo, J. W. Park, *et al.*, "Evaluating the utility of rapid point-of-care potassium testing for the early identification of hyperkalemia in patients with chronic kidney disease in the emergency department," *Yonsei medical journal,* vol. 55, pp. 1348-1353, 2014.
[261] A. Auvet, F. Espitalier, L. Grammatico-Guillon, M.-A. Nay, D. Elaroussi, M. Laffon, *et al.*, "Preanalytical conditions of point-of-care testing in the intensive care unit are decisive for analysis reliability," *Annals of intensive care,* vol. 6, pp. 57-57, 2016.
[262] P. Meybohm, K. Zacharowski, and C. F. Weber, "Point-of-care coagulation management in intensive care medicine," *Critical care (London, England),* vol. 17, pp. 218-218, 2013.
[263] O. M. Theusinger, P. Stein, and J. H. Levy, "Point of Care and Factor Concentrate-Based Coagulation Algorithms," *Transfusion Medicine and Hemotherapy,* vol. 42, pp. 115-121, 2015.
[264] C. Lazaridis, M. Yang, S. M. DeSantis, S. T. Luo, and C. S. Robertson, "Predictors of intensive care unit length of stay and intracranial pressure in severe traumatic brain injury," *Journal of critical care,* vol. 30, pp. 1258-1262, 2015.
[265] M. Hasani, M. A. Sahraian, M. Motamedi, and K. Mostehaghan, "Postmortem cerebrospinal fluid analysis in a general intensive care unit," *Indian Journal of Critical Care Medicine,* vol. 9, pp. 176-178, July 1, 2005 2005.
[266] F. A. Gers, D. Eck, and J. Schmidhuber, "Applying LSTM to Time Series Predictable Through Time-Window Approaches," in *Neural Nets WIRN Vietri-01: Proceedings of the 12th Italian Workshop on Neural Nets, Vietri sul Mare, Salerno, Italy, 17–19 May 2001,* R. Tagliaferri and M. Marinaro, Eds., ed London: Springer London, 2002, pp. 193-200.
[267] K. Kryszczuk and P. Hurley, "Estimation of the number of clusters using multiple clustering validity indices," in *International Workshop on Multiple Classifier Systems*, 2010, pp. 114-123.
[268] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, pp. 224-227, 1979.
[269] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *Journal of computational and applied mathematics,* vol. 20, pp. 53-65, 1987.
[270] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 63, pp. 411-423, 2001.
[271] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology,* vol. 48, pp. 229-240, 2011.
[272] S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, *et al.*, "Automatic classification of MR scans in Alzheimer's disease," *Brain,* vol. 131, pp. 681-689, 2008.

[273] (2016, 2/23). *Alzheimer's Facts and Figures*. Available: http://www.alz.org/facts/
[274] A. Alzheimer's, "2013 Alzheimer's disease facts and figures," *Alzheimer's & Dementia*, vol. 9, pp. 208-245, 3// 2013.
[275] B. Tejada-Vera and N. C. f. H. Statistics, *Mortality from Alzheimer's disease in the United States: Data for 2000 and 2010*: Citeseer, 2013.
[276] R. J. Perrin, A. M. Fagan, and D. M. Holtzman, "Multimodal techniques for diagnosis and prognosis of Alzheimer's disease," *Nature*, vol. 461, pp. 916-922, 2009.
[277] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, *et al.*, "Impact of the Alzheimer's Disease Neuroimaging Initiative, 2004 to 2014," *Alzheimer's & Dementia*, vol. 11, pp. 865-884, 7// 2015.
[278] T. Grimmer, C. Wutz, P. Alexopoulos, A. Drzezga, S. Förster, H. Förstl, *et al.*, "Visual versus fully-automated analyses of FDG-and Amyloid-PET for prediction of dementia due to Alzheimer's disease in mild cognitive impairment," *Journal of Nuclear Medicine*, p. jnumed. 115.163717, 2015.
[279] S. F. Eskildsen, P. Coupé, V. S. Fonov, J. C. Pruessner, D. L. Collins, and A. s. D. N. Initiative, "Structural imaging biomarkers of Alzheimer's disease: predicting disease progression," *Neurobiology of aging*, vol. 36, pp. S23-S31, 2015.
[280] K. Blennow, B. Dubois, A. M. Fagan, P. Lewczuk, M. J. de Leon, and H. Hampel, "Clinical utility of cerebrospinal fluid biomarkers in the diagnosis of early Alzheimer's disease," *Alzheimer's & Dementia*, vol. 11, pp. 58-69, 2015.
[281] R. Chaves, J. M. Gorriz, J. Ramirez, I. A. Illan, D. Salas-Gonzalez, and M. Gomez-Rio, "Efficient mining of association rules for the early diagnosis of Alzheimer's disease," *Phys Med Biol*, vol. 56, pp. 6047-63, Sep 21 2011.
[282] W. de Haan, Y. A. Pijnenburg, R. L. Strijers, Y. van der Made, W. M. van der Flier, P. Scheltens, *et al.*, "Functional neural network analysis in frontotemporal dementia and Alzheimer's disease using EEG and graph theory," *BMC Neurosci*, vol. 10, p. 101, 2009.
[283] B. T. Hyman, C. H. Phelps, T. G. Beach, E. H. Bigio, N. J. Cairns, M. C. Carrillo, *et al.*, "National Institute on Aging–Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease," *Alzheimer's & Dementia*, vol. 8, pp. 1-13, 2012.
[284] E. Mamikonyan, P. J. Moberg, A. Siderowf, J. E. Duda, T. Ten Have, H. I. Hurtig, *et al.*, "Mild cognitive impairment is common in Parkinson's disease patients with normal Mini-Mental State Examination (MMSE) scores," *Parkinsonism & related disorders*, vol. 15, pp. 226-231, 2009.
[285] J. Barnes, O. T. Carmichael, K. K. Leung, C. Schwarz, G. R. Ridgway, J. W. Bartlett, *et al.*, "Vascular and Alzheimer's disease markers independently predict brain atrophy rate in Alzheimer's Disease Neuroimaging Initiative controls," *Neurobiology of Aging*, vol. 34, pp. 1996-2002, 8// 2013.
[286] K. Henriksen, S. E. O'Bryant, H. Hampel, J. Q. Trojanowski, T. J. Montine, A. Jeromin, *et al.*, "The future of blood-based biomarkers for Alzheimer's disease," *Alzheimer's & Dementia*, vol. 10, pp. 115-131, 1// 2014.
[287] J. D. Doecke, S. M. Laws, N. G. Faux, and et al., "BLood-based protein biomarkers for diagnosis of alzheimer disease," *Archives of Neurology*, vol. 69, pp. 1318-1325, 2012.

[288] L. Glodzik, L. Mosconi, W. Tsui, S. de Santi, R. Zinkowski, E. Pirraglia, *et al.*, "Alzheimer's disease markers, hypertension, and gray matter damage in normal elderly," *Neurobiology of Aging,* vol. 33, pp. 1215-1227, 7// 2012.
[289] B. Dubois, H. H. Feldman, C. Jacova, H. Hampel, J. L. Molinuevo, K. Blennow, *et al.*, "Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria," *The Lancet Neurology,* vol. 13, pp. 614-629, 6// 2014.
[290] H. Geekiyanage, G. A. Jicha, P. T. Nelson, and C. Chan, "Blood serum miRNA: Non-invasive biomarkers for Alzheimer's disease," *Experimental Neurology,* vol. 235, pp. 491-496, 6// 2012.
[291] M. Dyrba, M. Grothe, T. Kirste, and S. J. Teipel, "Multimodal analysis of functional and structural disconnection in Alzheimer's disease using multiple kernel SVM," *Human Brain Mapping,* vol. 36, pp. 2118-2131, 2015.
[292] J. L. Shaffer, J. R. Petrella, F. C. Sheldon, K. R. Choudhury, V. D. Calhoun, R. E. Coleman, *et al.*, "Predicting Cognitive Decline in Subjects at Risk for Alzheimer Disease by Using Combined Cerebrospinal Fluid, MR Imaging, and PET Biomarkers," *Radiology,* vol. 266, pp. 583-591, 2013/02/01 2013.
[293] Z. Dai, C. Yan, Z. Wang, J. Wang, M. Xia, K. Li, *et al.*, "Discriminative analysis of early Alzheimer's disease using multi-modal imaging and multi-level characterization with multi-classifier (M3)," *NeuroImage,* vol. 59, pp. 2187-2195, 2/1/ 2012.
[294] M. Dyrba, F. Barkhof, A. Fellgiebel, M. Filippi, L. Hausner, K. Hauenstein, *et al.*, "Predicting Prodromal Alzheimer's Disease in Subjects with Mild Cognitive Impairment Using Machine Learning Classification of Multimodal Multicenter Diffusion-Tensor and Magnetic Resonance Imaging Data," *Journal of Neuroimaging,* vol. 25, pp. 738-747, 2015.
[295] M. Lorenzi, I. J. Simpson, A. F. Mendelson, S. B. Vos, M. J. Cardoso, M. Modat, *et al.*, "Multimodal Image Analysis in Alzheimer's Disease via Statistical Modelling of Non-local Intensity Correlations," *Scientific Reports,* vol. 6, p. 22161, 04/11
06/29/received
02/04/accepted 2016.
[296] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based similarity measures for multi-modal classification of Alzheimer's disease," *NeuroImage,* vol. 65, pp. 167-175, 1/15/ 2013.
[297] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage,* vol. 55, pp. 856-867, 4/1/ 2011.
[298] H. Wang, F. Nie, H. Huang, S. L. Risacher, A. J. Saykin, L. Shen, *et al.*, "Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning," *Bioinformatics,* vol. 28, pp. i127-i136, 2012.
[299] Y. Zhang, Z. Dong, P. Phillips, S. Wang, G. Ji, J. Yang, *et al.*, "Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning," *Frontiers in Computational Neuroscience,* vol. 9, p. 66, 2015.

[300] C. Qiao, D.-D. Lin, S.-L. Cao, and Y.-P. Wang, "The effective diagnosis of schizophrenia by using multi-layer RBMs deep networks," in *Bioinformatics and Biomedicine (BIBM), 2015 IEEE International Conference on*, 2015, pp. 603-606.
[301] H.-I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*, ed: Springer, 2013, pp. 583-590.
[302] S. Liu, S. Liu, W. Cai, H. Che, S. Pujol, R. Kikinis, *et al.*, "Multimodal Neuroimaging Feature Learning for Multiclass Diagnosis of Alzheimer's Disease," *Biomedical Engineering, IEEE Transactions on,* vol. 62, pp. 1132-1140, 2015.
[303] H.-I. Suk, S.-W. Lee, D. Shen, and A. S. D. N. Initiative, "Deep sparse multi-task learning for feature selection in Alzheimer's disease diagnosis," *Brain Structure and Function,* pp. 1-19, 2015.
[304] P. Schulam, F. Wigley, and S. Saria, "Clustering Longitudinal Clinical Marker Trajectories from Electronic Health Data: Applications to Phenotyping and Endotype Discovery," in *AAAI*, 2015, pp. 2956-2964.
[305] H.-I. Suk and D. Shen, "Deep learning-based feature representation for AD/MCI classification," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 583-590.
[306] H.-I. Suk, S.-W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for AD/MCI diagnosis," *NeuroImage,* vol. 101, pp. 569-582, 11/1/ 2014.
[307] J. Zhou and O. G. Troyanskaya, "Predicting effects of noncoding variants with deep learning-based sequence model," *Nature methods,* vol. 12, pp. 931-934, 2015.
[308] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 689-696.
[309] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. R. Jack, W. Jagust, *et al.*, "Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI)," *Alzheimer's & Dementia,* vol. 1, pp. 55-66, 2005.
[310] S. B. Eickhoff, K. E. Stephan, H. Mohlberg, C. Grefkes, G. R. Fink, K. Amunts, *et al.*, "A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data," *Neuroimage,* vol. 25, pp. 1325-1335, 2005.
[311] R. Miotto, L. Li, B. A. Kidd, and J. T. Dudley, "Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records," *Scientific Reports,* vol. 6, p. 26094, 05/17
01/28/received
04/27/accepted 2016.
[312] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167,* 2015.
[313] M. D. Ritchie, E. R. Holzinger, R. Li, S. A. Pendergrass, and D. Kim, "Methods of integrating data to uncover genotype-phenotype interactions," *Nature Reviews Genetics,* vol. 16, pp. 85-97, 2015.

[314] H. Mhaskar, Q. Liao, and T. Poggio, "Learning functions: when is deep better than shallow," *arXiv preprint arXiv:1603.00988,* 2016.
[315] K. Pasupa and W. Sunhem, "A comparison between shallow and deep architecture classifiers on small dataset," in *Information Technology and Electrical Engineering (ICITEE), 2016 8th International Conference on*, 2016, pp. 1-6.
[316] R. E. Hampson, D. Song, I. Opris, L. M. Santos, D. C. Shin, G. A. Gerhardt, *et al.*, "Facilitation of memory encoding in primate hippocampus by a neuroprosthesis that promotes task-specific neural firing," *Journal of neural engineering,* vol. 10, p. 066013, 2013.
[317] E. Zahedi, J. Dargahi, M. Kia, and M. Zadeh, "Gesture-Based Adaptive Haptic Guidance: A Comparison of Discriminative and Generative Modeling Approaches," *IEEE Robotics and Automation Letters,* vol. 2, pp. 1015-1022, 2017.
[318] L. Mescheder, S. Nowozin, and A. Geiger, "Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks," *arXiv preprint arXiv:1701.04722,* 2017.
[319] D. Yogatama, C. Dyer, W. Ling, and P. Blunsom, "Generative and discriminative text classification with recurrent neural networks," *arXiv preprint arXiv:1703.01898,* 2017.
[320] J. C. Mandel, D. A. Kreda, K. D. Mandl, I. S. Kohane, and R. B. Ramoni, "SMART on FHIR: a standards-based, interoperable apps platform for electronic health records," *Journal of the American Medical Informatics Association,* p. ocv189, 2016.
[321] C. G. Chute, J. Pathak, G. K. Savova, K. R. Bailey, M. I. Schor, L. A. Hart, *et al.*, "The SHARPn project on secondary use of Electronic Medical Record data: progress, plans, and possibilities," in *AMIA Annu Symp Proc*, 2011, pp. 248-56.
[322] K. Marsolo and S. A. Spooner, "Clinical genomics in the world of the electronic health record," *Genet Med,* vol. 15, pp. 786-91, Oct 2013.
[323] E. A. Crabtree, E. Brennan, A. Davis, and J. E. Squires, "Connecting Education to Quality: Engaging Medical Students in the Development of Evidence-Based Clinical Decision Support Tools," *Academic Medicine,* vol. 92, pp. 83-86, 2017.
[324] A. Belard, T. Buchman, J. Forsberg, B. K. Potter, C. J. Dente, A. Kirk, *et al.*, "Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care," *Journal of Clinical Monitoring and Computing,* vol. 31, pp. 261-271, 2017// 2017.
[325] K. Karnik, "FDA regulation of clinical decision support software," *Journal of Law and the Biosciences,* vol. 1, pp. 202-208, 2014.
[326] R. B. Myers, S. L. Jones, and D. F. Sittig, "Review of reported clinical information system adverse events in US Food and Drug Administration databases," *Appl Clin Inform,* vol. 2, pp. 63-74, 2011.
[327] !!! INVALID CITATION !!! [133, 134].
[328] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering,* vol. 22, pp. 1345-1359, 2010.
[329] Z. Shi, P. Siva, and T. Xiang, "Transfer learning by ranking for weakly supervised object annotation," *arXiv preprint arXiv:1705.00873,* 2017.

VITA

JANANI VENUGOPALAN

VENUGOPALAN was born in Chennai, India. She attended several public schools in India, received a B.Tech. in Biomedical Engineering, B.E. in Computer Science from Satyabhama University, India, in 2008 and a M.Tech. in Clinical Engineering from Indian Institute of Technology, Madras, India, Tennessee in 2010 before coming to Georgia Tech to pursue a doctorate in Biomedical Engineering. When she is not working on his research, Ms.Venugopalan enjoys reading fiction, watching movies, gaming with friends, hiking and martial arts.